Communications in
**Mathematical
Physics**

Continuous-Variable Nonlocality and Contextuality

Rui Soares Barbosa[1]**, Tom Douce**[2]**, Pierre-Emmanuel Emeriau**[3] ,
Elham Kashefi[3,4]**, Shane Mansfield**[5]

[1] INL – International Iberian Nanotechnology Laboratory, Braga, Portugal.
   E-mail: rui.soaresbarbosa@inl.int
[2] Paris, France. E-mail: doucetom@hotmail.fr
[3] Laboratoire d'Informatique de Paris 6, CNRS and Sorbonne Université, Paris, France.
   E-mail: pierre-emmanuel.emeriau@lip6.fr
[4] School of Informatics, University of Edinburgh, Edinburgh, UK.
   E-mail: elham.kashefi@lip6.fr
[5] Quandela, Massy, France. E-mail: shane.mansfield@quandela.com

Abstract: Contextuality is a non-classical behaviour that can be exhibited by quantum systems. It is increasingly studied for its relationship to quantum-over-classical advantages in informatic tasks. To date, it has largely been studied in discrete-variable scenarios, where observables take values in discrete and usually finite sets. Practically, on the other hand, continuous-variable scenarios offer some of the most promising candidates for implementing quantum computations and informatic protocols. Here we set out a framework for treating contextuality in continuous-variable scenarios. It is shown that the Fine–Abramsky–Brandenburger theorem extends to this setting, an important consequence of which is that Bell nonlocality can be viewed as a special case of contextuality, as in the discrete case. The contextual fraction, a quantifiable measure of contextuality that bears a precise relationship to Bell inequality violations and quantum advantages, is also defined in this setting. It is shown to be a non-increasing monotone with respect to classical operations that include binning to discretise data. Finally, we consider how the contextual fraction can be formulated as an infinite linear program. Through Lasserre relaxations, we are able to express this infinite linear program as a hierarchy of semi-definite programs that allow to calculate the contextual fraction with increasing accuracy.

Introduction

Contextuality is one of the principal markers of non-classical behaviour that can be exhibited by quantum systems. The Heisenberg uncertainty principle identified that certain pairs of quantum observables are incompatible, e.g. position and momentum. In operational terms, observing one will disturb the outcome statistics of the other. This is sometimes cited as evidence that not all observables can simultaneously be assigned definite values. Taking the mathematical formalism of quantum mechanics at face value, that is indeed the case, in stark contrast with classical physical theories. However, one may wonder whether it is possible to build a (presumably more fundamental) theory

more in accordance with our classical intuitions, but which still *matches* the empirical predictions of quantum mechanics. Put briefly, the fundamental question is then whether such quantum oddities are a necessary property of any theory that accurately describes nature, and thus have *empirical content*, or mere artifices of the mathematical formalism of quantum theory.

This question might be answered by attempting to build a hidden-variable model reproducing quantum-mechanical empirical predictions but with the further assumption that it be *noncontextual*. Roughly speaking, the latter imposes that the model must respect the basic assumptions that (i) hidden variables assign definite values to all the observable properties, and (ii) jointly performing *compatible* observables does not disturb the hidden variable. That these apparently simple assumptions are at odds with the empirical predictions of quantum mechanics is the content of the seminal theorems by Bell [19] and by Kochen & Specker [52].

Separately to its foundational importance, contextuality also has a more practical significance. A major application of quantum theory today is in quantum information and computation. There, one is primarily interested in what can be *done* with quantum systems and is beyond the capabilities of any classical implementation. So one is interested in the properties of the *correlations* realisable by quantum systems when compared to the kind of correlations that could arise from any classical theory. In this sense, aside from whatever foundational or physical significance one may wish (or not) to ascribe to contextuality, it has an undeniable practical significance in relation to quantum information and computation. In particular, it has now been identified as the essential ingredient for enabling a range of quantum-over-classical advantages in informatic tasks, which include the onset of universal quantum computing in certain computational models [6,7,20,46,74].

It is notable that to date the study of contextuality has largely focused on discrete variable scenarios and that the main frameworks and approaches to contextuality are tailored to modelling these, e.g. [8,13,26,33]. In such scenarios, observables can only take values in discrete, and usually finite, sets. Discrete variable scenarios typically arise in finite-dimensional quantum mechanics, e.g. when dealing with quantum registers in the form of systems of multiple qubits as is common in quantum information and computation theory.

Yet, from a practical perspective, continuous-variable quantum systems are emerging as some of the most promising candidates for implementing quantum informational and computational tasks [25,84]. The main reason for this is that they offer unrivalled possibilities for deterministic generation of large-scale resource states [86] and for highly-efficient measurements of certain observables. Together these cover many of the basic operations required in the one-way or measurement-based model of quantum computing [76], for example. Typical implementations are in optical systems where the continuous variables correspond to the position-like and momentum-like quadratures of the quantised modes of an electromagnetic field. Indeed position and momentum, as mentioned previously in relation to the uncertainty principle, are the prototypical examples of continuous variables in quantum mechanics.

Since quantum mechanics itself is infinite dimensional, it also makes sense from a foundational perspective to extend analyses of the key concept of contextuality to the continuous-variable setting. Furthermore, continuous variables can be useful when dealing with iteration, even when attention is restricted to finite-variable actions at discrete time steps, as is traditional in informatics. An interesting question, for example, is

whether contextuality arises and is of interest in such situations as the infinite behaviour of quantum random walks.

The main contributions of this article are the following:

- We present a robust framework for contextuality in continuous-variable scenarios that follows along the lines of the discrete-variable framework introduced by Abramsky and Brandenburger [8] (Sect. 3). We thus generalise this framework to deal with outcomes being valued on general measurable spaces, as well as to arbitrary (infinite) sets of measurement labels.

- We show that the Fine–Abramsky–Brandenburger theorem [8,36] extends to continuous variables (Sect. 4). This establishes that noncontextuality of an empirical behaviour, originally characterised by the existence of a deterministic hidden-variable model [19,52], can equivalently be characterised by the existence of a factorisable hidden-variable model, and that ultimately both of these are subsumed by a canonical form of hidden-variable model—a *global section* in the sheaf-theoretic perspective. An important consequence is that Bell nonlocality may be viewed as a special case of contextuality in continuous-variable scenarios just as for discrete-variable scenarios.

- The contextual fraction, a quantifiable measure of contextuality that bears a precise relationship to Bell inequality violations and quantum advantages [6], can also be defined in this setting using infinite linear programming (Sect. 5). It is shown to be a non-increasing monotone with respect to the free operations of a resource theory for contextuality [4,6]. Crucially, these include the common operation of binning to discretise data. A consequence is that any witness of contextuality on discretised empirical data also witnesses and gives a lower bound on genuine continuous-variable contextuality.

- While the infinite linear programs are of theoretical importance and capture exactly the quantity and Bell-like inequalities in which we are interested, they are not directly useful for actual numerical computations. To get around this limitation, we introduce a hierarchy of semi-definite programs which are relaxations of the original problem and whose values converge monotonically to the contextual fraction (Sect. 8). This applies in the restricted setting where there is a finite set of measurement labels.

Related work. Note that we are specifically interested in scenarios involving observables with continuous spectra, or in more operational language, measurements with continuous outcome spaces. We still consider scenarios featuring only discrete sets of observables or measurements, as is typical in continuous-variable quantum computing. The possibility of considering contextuality in settings with continuous measurement spaces has also been evoked in [30]. We also note that several prior works have explicitly considered contextuality in continuous-variable systems [14,42,50,57,64,71,82]. Our approach is distinct from these in that it provides a genuinely continuous-variable treatment of contextuality itself as opposed to embedding discrete-variable contextuality arguments into, or extracting them from, continuous-variable systems.

1. Continuous-Variable Behaviours

In this section we provide a brief motivational example for the kind of continuous-variable empirical behaviour we are interested in analysing. The approach applies generally to any hypothetical empirical data, including those that do not admit a quantum realisation (e.g. the PR box from Ref. [72]). But also, in particular, it does of course apply to empirical data arising from quantum mechanics, in that the statistics arise from a state and measurements on a quantum system according to the Born rule. Indeed,
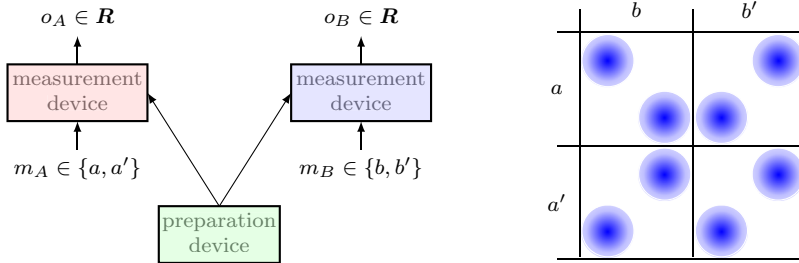
Fig. 1. [Left] operational depiction of a typical bipartite experimental scenario. [Right] Hypothetical probability density plots for empirical data arising from such an experiment. Cf. the discrete-variable probability tables of [60,62]

quantum theory provides the main motivation for this study and more broadly for the sheaf-theoretic approach, because of a feature that may arise in empirical models having quantum but not classical realisations: which we refer to as *contextuality*.

Suppose that we can interact with a system by performing measurements on it and observing their outcomes. A feature of quantum systems is that not all observables commute, so that certain combinations of measurements are incompatible.

At best, we can obtain empirical observational data for contexts in which only compatible measurements are performed, which can be collected by running the experiment repeatedly. As we shall make more precise in Sects. 3 and 4, contextuality arises when the empirical data obtained is inconsistent with the assumption that for each run of the experiment the system has a global and context-independent assignment of values to all of its observable properties.

To take an operational perspective, a typical example of an experimental setup or scenario that we consider is the one depicted in Fig. 1 [left]. In this scenario, a system is prepared in some fixed bipartite state, following which parties A and B may each choose between two measurement settings, $m_A \in \{a, a'\}$ for A and $m_B \in \{b, b'\}$ for B. We assume that outcomes of each measurement live in R, which typically will be a bounded measurable subspace of the real numbers $\mathbb{R}$ (with its Borel σ-algebra). Depending on which choices of inputs were made, the empirical data might for example be distributed according to one of the four hypothetical probability density plots in R^2 depicted in Fig. 1 [right]. This scenario and hypothetical empirical behaviour has been considered elsewhere [50] as a continuous-variable version of the Popescu–Rohrlich (PR) box [72].

2. Preliminaries on Measures and Probability

In order to properly treat probability on continuous-variable spaces, it is necessary to introduce a modicum of measure theory. This section serves to recall some basic ideas and to fix notation. The reader may choose to skip the section and consult it as reference for the remainder of the article.

A *measurable space* is a pair $X = \langle X, \mathcal{F} \rangle$ consisting of a set X and a σ-algebra (or σ-field) $\mathcal{F}$ on X, i.e. a family of subsets of X containing the empty set and closed under complementation and countable unions. In some sense, this specifies the subsets of X that can be assigned a 'size', and which are therefore called the *measurable sets* of X. Throughout this paper, we follow the convention of using boldface to denote the measurable space and the same symbol in regular face for its underlying set.

A trivial example of a σ-algebra over any set X is its powerset $\mathcal{P}(X)$, which gives the discrete measurable space $\langle X, \mathcal{P}(X)\rangle$, where every set is measurable. This is typically used when X is countable (finite or countably infinite), in which case this discrete σ-algebra is generated by the singletons. Another example, of central importance in measure theory, is $\langle \mathbb{R}, \mathcal{B}_{\mathbb{R}}\rangle$, where $\mathcal{B}_{\mathbb{R}}$ is the σ-algebra generated from the open sets of $\mathbb{R}$, whose elements are called the Borel sets. Working with Borel sets avoids the problems that would arise if we naively attempted to measure or assign probabilities to points in the continuum. More generally, any topological space gives rise to a Borel measurable space in this fashion.

A *measurable function* between measurable spaces $X = \langle X, \mathcal{F}_X\rangle$ and $Y = \langle Y, \mathcal{F}_Y\rangle$ is a function $f\colon X \longrightarrow Y$ between the underlying sets whose preimage preserves measurable sets, i.e. such that, for any $E \in \mathcal{F}_Y$, $f^{-1}(E) \in \mathcal{F}_X$. This is analogous to the definition of a continuous function between topological spaces. Clearly, the identity function is measurable and measurable functions compose. We will denote by Meas the category whose objects are measurable spaces and whose morphisms are measurable functions.

The *product* of two measurable spaces $X_1 = \langle X_1, \mathcal{F}_1\rangle$ and $X_2 = \langle X_2, \mathcal{F}_2\rangle$ is the measurable space

$$X_1 \times X_2 = \langle X_1 \times X_2, \mathcal{F}_1 \otimes \mathcal{F}_2\rangle , \tag{1}$$

where the Cartesian product of the underlying sets, $X_1 \times X_2$, is equipped with the so-called tensor product σ-algebra $\mathcal{F}_1 \otimes \mathcal{F}_2$, which is the σ-algebra generated by the 'rectangles', subsets of the form $E_1 \times E_2$ with $E_1 \in \mathcal{F}_1$ and $E_2 \in \mathcal{F}_2$. This is the categorical (binary) product in Meas.

We shall also need to deal with *infinite products* of measurable spaces. The generalisation is analogous to that for products of topological spaces, where the box topology (generated by 'rectangles') is no longer the most natural choice when dealing with infinite families, but rather the topology generated by 'cylinders'. Let I be an arbitrary index set. The product of measurable spaces $(X_i = \langle X_i, \mathcal{F}_i\rangle)_{i \in I}$ is the measurable space

$$\prod_{i \in I} X_i = \left\langle \prod_{i \in I} X_i, \bigotimes_{i \in I} \mathcal{F}_i \right\rangle , \tag{2}$$

where $X_I = \prod_{i \in I} X_i$ is the Cartesian product of the underlying sets, and $\mathcal{F}_I = \bigotimes_{i \in I} \mathcal{F}_i$ is the σ-algebra generated by subsets of $\prod_{i \in I} X_i$ of the form $\prod_{i \in I} E_i$ where $E_i \subseteq X_i$ for all $i \in I$ and $E_i \neq X_i$ for only finitely many $i \in I$. This is the smallest σ-algebra that makes the projection maps $\pi_k\colon \prod_{i \in I} X_i \longrightarrow X_k$ measurable. It therefore corresponds to the categorical (arbitrary) product in Meas.

A *measure* on a measurable space $X = \langle X, \mathcal{F}\rangle$ is a function $\mu\colon \mathcal{F} \longrightarrow \overline{\mathbb{R}}$ from the σ-algebra to the extended real numbers $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ satisfying:

(i) [nonnegativity] $\mu(E) \geq 0$ for all $E \in \mathcal{F}$;
(ii) [null empty set] $\mu(\emptyset) = 0$;
(iii) [σ-additivity] for any countable family $(E_i)_{i=1}^{\infty}$ of pairwise disjoint measurable sets, it holds that $\mu(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} \mu(E_i)$.

A measure on X allows one to *integrate* well-behaved[1] measurable functions $f\colon X \longrightarrow \langle \mathbb{R}, \mathcal{B}_{\mathbb{R}}\rangle$ to obtain a real value, denoted $\int_X f \ \mathrm{d}\mu$ or $\int_{x \in X} f(x) \ \mathrm{d}\mu(x)$. The simplest

---

[1] For a comprehensive treatment, we refer the reader to e.g. [21], or to [68] for a beautiful and more concise introduction aimed particularly at computer scientists.

example of such a measurable function is the *indicator function* of a measurable set $E \in \mathcal{F}$:

$$\chi_E(x) := \begin{cases} 1 & \text{if } x \in E \\ 0 & \text{if } x \notin E. \end{cases}$$

For any measure μ on X, its integral yields

$$\int_X \chi_E \ \mathrm{d}\mu = \mu(E) \, . \tag{3}$$

A measure μ is finite if $\mu(X) < \infty$ and in particular it is a *probability measure* if $\mu(X) = 1$. We will denote by $\mathbb{M}(X)$ and $\mathbb{P}(X)$, respectively, the sets of measures and probability measures on the measurable space X.

A measurable function $f \colon X \longrightarrow Y$ carries any measure μ on X to a measure $f_*\mu$ on Y. This *push-forward* measure is given by $f_*\mu(E) = \mu(f^{-1}(E))$ for any set E measurable in Y. An important use of push-forward measures is that for any integrable function $g \colon Y \longrightarrow \langle \mathbb{R}, \mathcal{B}_{\mathbb{R}} \rangle$, it allows us to write the following change-of-variables formula

$$\int_Y g \ \mathrm{d} f_*\mu = \int_X g \circ f \ \mathrm{d}\mu \, . \tag{4}$$

The push-forward operation preserves the total measure, hence it takes $\mathbb{P}(X)$ to $\mathbb{P}(Y)$.

A case that will be of particular interest to us is the push-forward of a measure μ on a product space $X_1 \times X_2$ along a projection $\pi_i \colon X_1 \times X_2 \longrightarrow X_i$: this yields the *marginal measure* $\mu|_{X_i} = \pi_{i*}\mu$, where e.g. for E measurable in X_1, $\mu|_{X_1}(E) = \mu(\pi_1^{-1}(E)) = \mu(E \times X_2)$.

In the opposite direction, given measures μ_1 on X_1 and μ_2 on X_2, a *product measure* $\mu_1 \times \mu_2$ is a measure on the product measurable space $X_1 \times X_2$ satisfying $(\mu_1 \times \mu_2)(E_1 \times E_2) = \mu_1(E_1)\mu_2(E_2)$ for all $E_1 \in \mathcal{F}_1$ and $E_2 \in \mathcal{F}_2$. For probability measures, there is a uniquely determined product measure.[2] The analogous, much more general statement also holds for arbitrary products of probability measures (see e.g. [83, section 11.2]).

We can view $\mathbb{M}$ as a map that takes a measurable space to the set of measures on that space, and similarly for $\mathbb{P}$. These become functors $\mathsf{Meas} \longrightarrow \mathsf{Set}$ if we define the action on morphisms to be the push-forward operation. Explicitly we set $\mathbb{M}(f) := f_* \colon \mathbb{M}(X) \longrightarrow \mathbb{M}(Y)::\mu \longmapsto f_*\mu$, where $f \colon X \longrightarrow Y$ is a measurable function, and similarly for $\mathbb{P}$.

Remarkably, the set $\mathbb{P}(X)$ of probability measures on X can itself be made into a measurable space by equipping it with the least σ-algebra that makes the evaluation functions

$$\mathsf{ev}_E \colon \mathbb{P}(X) \longrightarrow [0, 1]::\mu \longmapsto \mu(E)$$

measurable for all $E \in \mathcal{F}_X$.[3] This yields an endofunctor $\mathbb{P} \colon \mathsf{Meas} \longrightarrow \mathsf{Meas}$, which moreover has the structure of a monad, called the Giry monad [39]. The unit of this monad is given by

$$\eta_X \colon X \longrightarrow \mathbb{P}(X)::x \longmapsto \delta_x$$

---

[2] In fact, this holds more generally for σ-finite measures, i.e. when the space is a countable union of sets of finite measure.

[3] More concretely, it is the σ-algebra generated by the sets $\mathsf{ev}_E^{-1}([0, r)) = \{\mu \in \mathbb{P}(X) \mid \mu(E) < r\}$ with $E \in \mathcal{F}_X$ and $r \in [0, 1]$.

where δ_x is the *Dirac measure*, or point mass, at x given by $\delta_x(E) := \chi_E(x)$. Multiplication of the monad is given by

$$\mu_X \colon \mathbb{P}(\mathbb{P}(X)) \longrightarrow \mathbb{P}(X)$$

which takes a probability measure P on $\mathbb{P}(X)$ to its 'average', a probability measure $\mu_X(P)$ on X, $\mu_X(P) \colon \mathcal{F}_X \longrightarrow [0, 1]$, whose value on a measurable set $E \in \mathcal{F}_X$ is given by $\mu_X(P)(E) := \int_{\mathbb{P}(X)} \mathsf{ev}_E \, \mathrm{d} P$.

The Kleisli category of this monad is the category of Markov kernels, which represent continuous-variable probabilistic maps and generalise the discrete notion of stochastic matrix. Concretely, a *Markov kernel* between measurable spaces $X = \langle X, \mathcal{F}_X \rangle$ and $Y = \langle Y, \mathcal{F}_Y \rangle$ is a function $k \colon X \times \mathcal{F}_Y \longrightarrow [0, 1]$ such that:

(i) for all $E \in \mathcal{F}_Y$, $k(-, E) \colon X \longrightarrow [0, 1]$ is a measurable function;[4]
(ii) for all $x \in X$, $k(x, -) \colon \mathcal{F}_Y \longrightarrow [0, 1]$ is a probability measure.

3. Framework

In this section, we follow closely the discrete-variable framework of [8] in more formally describing the kinds of experimental scenarios in which we are interested and the empirical behaviours that arise on these, although some extra care is required for dealing with continuous variables.

Measurement scenarios

Definition 1. A *measurement scenario* is a triple $\langle X, \mathcal{M}, \boldsymbol{O} \rangle$ whose elements are specified as follows.

– X is a (possibly infinite) set of *measurement labels*.
– $\mathcal{M}$ is a covering family of subsets of X, i.e. such that $\bigcup \mathcal{M} = X$. The elements $C \in \mathcal{M}$ are called *maximal contexts* and represent maximal sets of compatible observables. We therefore require that $\mathcal{M}$ be an anti-chain with respect to subset inclusion, i.e. that no element of this family is a proper subset of another. Any subset of a maximal context also represents a set of compatible measurements, and we refer to elements of $\mathcal{U} := \{U \subseteq C \mid C \in \mathcal{M}\}$ as *contexts*.[5]
– $\boldsymbol{O} = (\boldsymbol{O}_x)_{x \in X}$ specifies a measurable *space of outcomes* $\boldsymbol{O}_x = \langle O_x, \mathcal{F}_x \rangle$ for each measurement $x \in X$.

Measurement scenarios can be understood as providing a concise description of the kind of experimental setup that is being considered. For example, the setup represented in Fig. 1 is described by the measurement scenario:

$$X = \{a, a', b, b'\}, \qquad \mathcal{M} = \{\{a, b\}, \{a, b'\}, \{a', b\}, \{a', b'\}\}, \qquad \boldsymbol{O}_x = \boldsymbol{R}, \quad (5)$$

where $\boldsymbol{R}$ is a bounded measurable subspace of $\langle \mathbb{R}, \mathcal{B}_{\mathbb{R}} \rangle$.

---

[4] The space $[0, 1]$ is assumed to be equipped with its Borel σ-algebra.

[5] While it is more convenient to specify $\mathcal{M}$, note that the set of contexts $\mathcal{U}$ carries exactly the same information. It forms an abstract simplicial complex whose simplices are the contexts and whose facets are the maximal context. This combinatorial topological structure is emphasised in some presentations [4,15,16, 29,49].

If some set of measurements $U \subseteq X$ is considered together, there is a joint outcome space given by the product of the respective outcome spaces (see Eq. (2)),

$$\boldsymbol{O}_U := \prod_{x \in U} \boldsymbol{O}_x = \langle O_U, \mathcal{F}_U \rangle = \left\langle \prod_{x \in U} O_x \,, \, \bigotimes_{x \in U} \mathcal{F}_x \right\rangle .$$

The map $\mathcal{E}$ that maps $U \subseteq X$ to $\mathcal{E}(U) = \boldsymbol{O}_U$ is called the *event sheaf* as concretely it assigns to any set of measurements information about the outcome events that could result from jointly performing them. Note that as well as applying the map to valid contexts $U \in \mathcal{U}$ we will see that it can also be of interest to consider hypothetical outcome spaces for sets of measurements that do not necessarily form valid contexts, in particular $\mathcal{E}(X) = \boldsymbol{O}_X$, the joint outcome space for *all* measurements. Moreover, as we will briefly discuss, this map satisfies the conditions to be a sheaf $\mathcal{E} : \mathcal{P}(X)^{\mathsf{op}} \longrightarrow \mathsf{Meas}$, where $\mathcal{P}(X)$ denotes the powerset of X, similarly to its discrete-variable analogue in [8].

The language of sheaves

Sheaves are widely used in modern mathematics. They might roughly be thought of as providing a means of assigning information to the open sets of a topological space in such a way that information can be restricted to smaller open sets and consistent information on a family of open sets can be uniquely 'glued' on their union.[6] In this work we are concerned with discrete topological spaces whose points represent measurements, and the information that we are interested in assigning has to do with outcome spaces for these measurements and probability measures on these outcome spaces. Sheaves can be defined concisely in category-theoretic terms as contravariant functors (presheaves) satisfying an additional gluing condition, though in what follows we will also give a more concrete description in terms of restriction maps. Categorically, the event sheaf is a functor $\mathcal{E} : \mathcal{P}(X)^{\mathsf{op}} \longrightarrow \mathsf{Meas}$ where $\mathcal{P}(X)$ is viewed as a category in the standard way for partial orders, with morphisms corresponding to subset inclusions.

Sheaves come with a notion of *restriction*. In our example, restriction arises in the following way: whenever $U, V \in \mathcal{P}(X)$ with $U \subseteq V$ we have an obvious restriction map $\rho_U^V : \mathcal{E}(V) \longrightarrow \mathcal{E}(U)$ which simply projects from the product outcome space for V to that for U. Note that ρ_U^U is the identity map for any $U \in \mathcal{P}(X)$ and that if $U \subseteq V \subseteq W$ in $\mathcal{P}(X)$ then $\rho_U^V \circ \rho_V^W = \rho_U^W$. Already this is enough to show that $\mathcal{E}$ is a *presheaf*. In categorical terms it establishes functoriality. Our map assigns outcome spaces $\mathcal{E}(U) = \boldsymbol{O}_U$ to sets of measurements $U \in \mathcal{P}(X)$, and in sheaf-theoretic terminology elements of these outcome spaces are called *sections* over U. Sections over X are called *global sections*. For an inclusion $U \subseteq V$ and a section $\mathbf{o} \in \mathcal{E}(V) = O_V$, it is often more convenient to use the notation $\mathbf{o}|_U$ to denote $\rho_U^V(\mathbf{o}) \in \mathcal{E}(U) = O_U$, the restriction of $\mathbf{o}$ to U.

Additionally, the unique gluing property holds for $\mathcal{E}$. Suppose that $\mathcal{N} \subseteq \mathcal{P}(X)$ and we have an $\mathcal{N}$-indexed family of sections $(\mathbf{o}_U \in O_U)_{U \in \mathcal{N}}$ that is compatible in the sense that its elements agree on overlaps, i.e. that for all $U, V \in \mathcal{N}$, $\mathbf{o}_U|_{U \cap V} = \mathbf{o}_V|_{U \cap V}$. Then these sections can always be 'glued' together in a unique fashion to obtain a section $\mathbf{o}_N$ over $N := \cup \mathcal{N}$ such that $\mathbf{o}_N|_U = \mathbf{o}_U$ for all $U \in \mathcal{N}$. This makes $\mathcal{E}$ a *sheaf*.

---

[6] For a comprehensive reference on sheaf theory see e.g. [59].

We will primarily be concerned with probability measures on outcome spaces. For this, we recall that the Giry monad $\mathbb{P}\colon \mathsf{Meas} \longrightarrow \mathsf{Meas}$ takes a measurable space and returns the probability measures over that space. Composing it with the event sheaf yields the map $\mathbb{P}\circ\mathcal{E}$ that takes any context and returns the probability measures on its joint outcome space. In fact, this is a presheaf $\mathbb{P}\circ\mathcal{E}\colon \mathcal{P}(X)^{\mathsf{op}} \longrightarrow \mathsf{Meas}$, where restriction on sections is given by marginalisation of probability measures. Note that marginalisation simply corresponds to the push-forward of a measure along projections to a component of the product space, which are precisely the restriction maps of $\mathcal{E}$. Note, however, that this presheaf does not satisfy the gluing condition and thus it crucially is not a sheaf.

Empirical models

Definition 2. An *empirical model* on a measurement scenario $\langle X, \mathcal{M}, \boldsymbol{O}\rangle$ is a compatible family for the presheaf $\mathbb{P}\circ\mathcal{E}$ on the cover $\mathcal{M}$. Concretely, it is a family $e = (e_C)_{C\in\mathcal{M}}$, where e_C is a probability measure on the space $\mathcal{E}(C) = \boldsymbol{O}_C$ for each maximal context $C \in \mathcal{M}$, which satisfies the compatibility condition:

$$e_C|_{C\cap C'} = e_{C'}|_{C\cap C'} .$$

Empirical models capture in a precise way the probabilistic *behaviours* that may arise upon performing measurements on physical systems. The compatibility condition ensures that the empirical behaviour of a given measurement or compatible subset of measurements is independent of which other compatible measurements might be performed along with them. This is sometimes referred to as the *no-disturbance* condition. A special case is *no-signalling*, which applies in multi-party or Bell scenarios such as that of Fig. 1 and Eq. (5). In that case, contexts consist of measurements that are supposed to occur in space-like separated locations, and compatibility ensures for instance that the choice of performing a or a' at the first location does not affect the empirical behaviour at the second location, i.e. $e_{\{a,b\}}|_{\{b\}} = e_{\{a',b\}}|_{\{b\}}$.

Note also that while empirical models may arise from the predictions of quantum theory, their definition is theory-independent. This means that empirical models can just as well describe hypothetical behaviours beyond what can be achieved by quantum mechanics such as the well-studied Popescu–Rohrlich box [72]. This can be useful in probing the limits of quantum theory and in singling out what distinguishes and characterises quantum theory within larger spaces of probabilistic theories, both well-established lines of research in quantum foundations.

Sheaf-theoretically. An empirical model is a compatible family of sections for the presheaf $\mathbb{P}\circ\mathcal{E}$ indexed by the maximal contexts of the measurement scenario. A natural question that may occur at this point is whether these sections can be glued to form a global section, and this is what we address next.

Extendability and contextuality

Definition 3. An empirical model e on a scenario $\langle X, \mathcal{M}, \boldsymbol{O}\rangle$ is *extendable* (or *non-contextual*[7]) if there is a probability measure μ on the space $\mathcal{E}(X) = \boldsymbol{O}_X$ such that $\mu|_C = e_C$ for every $C \in \mathcal{M}$.[8]

Recall that $\boldsymbol{O}_X$ is the global outcome space, whose elements correspond to global assignments of outcomes to all the measurements in the given scenario. Of course, it is

---

[7] In the language of [40], (non)contextuality here—and throughout this article—refers to global (non)contextuality as opposed to Bell (non)contextuality.

[8] Notions of partial extendability have also been considered in the discrete setting in [61,79].

not always the case that X is a valid context, and if it were then $\mu = e_X$ would trivially extend the empirical model. The question of the existence of such a probability measure that recovers the context-wise empirical content of e is particularly significant. When it exists, it amounts to a way of modelling the observed behaviour as arising stochastically from the behaviours of underlying states, identified with the elements of O_X, each of which deterministically assigns outcomes to *all* the measurements in X independently of the measurement context that is actually performed. If an empirical model is not extendable it is said to be *contextual*. Furthermore, we will say that it is *Bell nonlocal* in the special setting of so-called Bell scenarios, where the compatibility structure of observables is obtained from space-like separation.

Sheaf-theoretically. A contextual empirical model is a compatible family of sections for the presheaf $\mathbb{P} \circ \mathcal{E}$ over the contexts of the measurement scenario that cannot be glued into a global section. Contextuality thus arises as the tension between local consistency and global inconsistency.

4. A FAB Theorem

Quantum theory presents a number of non-intuitive features. For instance, Einstein, Podolsky and Rosen (EPR) identified early on that if the quantum description of the world is taken as fundamental then entanglement poses a problem of "spooky action at a distance" [35]. Their conclusion was that quantum theory should be consistent with a deeper or more complete description of the physical world, in which such problems would disappear. The import of seminal foundational results like the Bell [18] and Bell–Kochen–Specker [19,52] theorems is that they identify such non-intuitive behaviours and then rule out the possibility of finding *any* underlying model for them that would not suffer from the same issues. Incidentally, we note that the EPR paradox was originally presented in terms of continuous variables, whereas Bell's theorem addressed a discrete variable analogue of it.

In the previous section, we characterised contextuality of an empirical model by the absence of a global section for that empirical model. We also saw that global sections capture one kind of underlying model for the behaviour, namely via deterministic global states that assign predefined outcomes to all measurements. This is precisely the kind of model referred to in the Kochen–Specker theorem [52]. Bell's theorem, on the other hand, pertains to a different kind of hidden-variable model, where the salient feature—Bell locality—is a kind of factorisability rather than determinism. Fine [36] showed that in one important measurement scenario (that of the concrete example from Fig. 1) the existence of one kind of model is equivalent to existence of the other. Abramsky and Brandenburger [8] proved in full generality that this existential equivalence holds for any discrete-variable measurement scenario, and that global sections of $\mathbb{P} \circ \mathcal{E}$ provide a canonical form of hidden-variable model.

In this section, we prove a Fine–Abramsky–Brandenburger theorem in the continuous-variable setting. It establishes that in this setting there is also an unambiguous, unified description of Bell locality and noncontextuality, which is captured in a canonical way by the notion of extendability.

We will begin by introducing hidden-variable models in a more precise way. The idea is that there exists some space Λ of hidden variables, which determine the empirical behaviour. However, elements of this space may not be directly empirically accessible themselves, so we allow that we might only have probabilistic information about them in the form of a probability measure p on Λ. The empirically observable behaviour should then arise as an average over the hidden-variable behaviours.

Definition 4. Let $\langle X, \mathcal{M}, \boldsymbol{O} \rangle$ be a measurement scenario. A *hidden-variable* model[9] on this scenario consists of the following ingredients:

  – A measurable space $\boldsymbol{\Lambda} = \langle \Lambda, \mathcal{F}_\Lambda \rangle$ of *hidden variables*.
  – A probability measure p on $\boldsymbol{\Lambda}$.
  – For each maximal context $C \in \mathcal{M}$, a probability kernel $k_C \colon \boldsymbol{\Lambda} \longrightarrow \mathcal{E}(C)$,[10] satis-
  fying the following compatibility condition: for any maximal contexts $C, C' \in \mathcal{M}$,

$$\forall \lambda \in \Lambda. \quad k_C(\lambda, -)|_{C \cap C'} = k_{C'}(\lambda, -)|_{C \cap C'} . \tag{6}$$

Remark 1. Equivalently, we can regard Eq. (6) as defining a function $\underline{k}$ from Λ to the set of empirical models over $\langle X, \mathcal{M}, \boldsymbol{O} \rangle$. The function assigns to each $\lambda \in \Lambda$ the empirical model $\underline{k}(\lambda) := (\underline{k}(\lambda)_C)_{C \in \mathcal{M}}$, where the correspondence with the definition above is via $\underline{k}(\lambda)_C = k_C(\lambda, -)$. This function must be 'measurable' in $\boldsymbol{\Lambda}$ in the sense that $\underline{k}(-)_C(B) \colon \Lambda \longrightarrow [0, 1]$ is a measurable function for all $C \in \mathcal{M}$ and $B \in \mathcal{F}_C$.

Definition 5. Let $\langle X, \mathcal{M}, \boldsymbol{O} \rangle$ be a measurement scenario and $\langle \boldsymbol{\Lambda}, p, k \rangle$ be a hidden-variable model. Then the corresponding empirical model e is given as follows: for any maximal context $C \in \mathcal{M}$ and measurable set of joint outcomes $B \in \mathcal{F}_C$,

$$e_C(B) = \int_\Lambda k_C(-, B) \, \mathrm{d} \, p = \int_{\lambda \in \Lambda} k_C(\lambda, B) \, \mathrm{d} \, p(\lambda) .$$

Note that our definition of hidden-variable model assumes the properties known as λ-independence [31] and parameter-independence [47,78]. The former corresponds to the fact that the probability measure p on the hidden-variable space is independent of the measurement context to be performed, while the latter corresponds to the compatibility condition (6), which also ensures that the corresponding empirical model satisfies no-signalling [23]. We refer the reader to [24] for a detailed discussion of these and other properties of hidden-variable models specifically in the case of multi-party Bell scenarios.

The idea behind the introduction of hidden variables is that they could *explain away* some of the more non-intuitive aspects of the empirical predictions of quantum mechanics, which would then arise as resulting from an incomplete knowledge of the true state of a system rather than being a fundamental feature. There is some precedent for this in physical theories: for instance, statistical mechanics—a probabilistic theory—admits a deeper, albeit usually unwieldily complex, description in terms of classical mechanics, which is purely deterministic. Therefore, it is desirable to impose conditions on hidden-variable models which amount to requiring that they behave in some sense *classically* when conditioned on each particular value of the hidden variable λ. This motivates the notions of deterministic and of factorisable hidden-variable models.

Definition 6. A hidden-variable model $\langle \boldsymbol{\Lambda}, p, k \rangle$ is *deterministic* if the probability kernel $k_C(\lambda, -) \colon \mathcal{F}_C \longrightarrow [0, 1]$ is a Dirac measure for every $\lambda \in \Lambda$ and for every maximal context $C \in \mathcal{M}$; in other words, there is an assignment $\boldsymbol{o} \in O_C$ such that $k_C(\lambda, -) = \delta_{\boldsymbol{o}}$.

---

[9] The alternative term *ontological model* [81] has become widely used in quantum foundations in recent years. It indicates that the hidden variable, sometimes referred to as the ontic state, is supposed to provide an underlying description of the physical world at perhaps a more fundamental level than the empirical-level description via the quantum state for example.

[10] Recall from Sect. 2 that a probability kernel $k_C \colon \boldsymbol{\Lambda} \longrightarrow \mathcal{E}(C)$ is a function $k_C \colon \Lambda \times \mathcal{F}_C \longrightarrow [0, 1]$ which is a measurable function in the first argument and a probability measure in the second argument.

In general discussions on hidden-variable models (e.g. [24]), the condition above, requiring that each hidden variable determines a unique joint outcome for each measurement context, is sometimes referred to as weak determinism. This is contraposed to strong determinism, which demands not only that each hidden variable fix a deterministic outcome to each individual measurement, but that this outcome be independent of the context in which the measurement is performed. Note, however, that since our definition of hidden-variable models assumes the compatilibity condition (6), i.e. parameter-independence, both notions of determinism coincide [23].

Definition 7. A hidden-variable model $\langle \Lambda, p, k \rangle$ is *factorisable* if $k_C(\lambda, -) \colon \mathcal{F}_C \longrightarrow [0, 1]$ factorises as a product measure for every $\lambda \in \Lambda$ and for every maximal context $C \in \mathcal{M}$. That is, for any family of measurable sets $(B_x \in \mathcal{F}_x)_{x \in C}$ with $B_x \neq O_x$ only for finitely many $x \in C$,

$$k_C\left(\lambda, \prod_{x \in C} B_x\right) = \prod_{x \in C} k_C|_{\{x\}}(\lambda, B_x)$$

where $k_C|_{\{x\}}(\lambda, -)$ is the marginal of the probability measure $k_C(\lambda, -)$ on $\boldsymbol{O}_C = \prod_{x \in C} \boldsymbol{O}_x$ to the space $\boldsymbol{O}_{\{x\}} = \boldsymbol{O}_x$.[11]

Remark 2. In other words, if we consider elements of Λ as inaccessible 'empirical' models—i.e. if we use the alternative definition of hidden-variable models using the map $\underline{k}$ (see Remark 1)—then factorisability is the requirement that each of these be factorisable in the sense that

$$\underline{k}_C(\lambda)\left(\prod_{x \in C} B_x\right) = \prod_{x \in C} \underline{k}_C(\lambda)|_{\{x\}}(B_x)$$

where $\underline{k}_C|_{\{x\}}(\lambda)$ is the marginal of the probability measure $\underline{k}_C(\lambda)$ on $\boldsymbol{O}_C = \prod_{x \in C} \boldsymbol{O}_x$ to the space $\boldsymbol{O}_x$.

We now prove the continuous-variable analogue of the theorem proved in the discrete probability setting by Abramsky and Brandenburger [8, Proposition 3.1 and Theorem 8.1], generalising the result of Fine [36] to arbitrary measurement scenarios.

In particular, this result shows that the measurable space $\mathcal{E}(X) = \boldsymbol{O}_X$ provides a canonical hidden-variable space. The proof that $(1) \Rightarrow (2)$ in the Theorem below shows how a global probability measure extending an empirical model e can be understood as giving a deterministic hidden-variable model with $\Lambda = \mathcal{E}(X)$. Canonicity is then established together with the proof that $(3) \Rightarrow (1)$, to the effect that if a given empirical model admits any factorisable hidden-variable model then it admits a deterministic model of the form just mentioned (with $\mathcal{E}(X)$ being the hidden-variable space).

Theorem 1. *Let e be an empirical model on a measurement scenario $\langle X, \mathcal{M}, \boldsymbol{O} \rangle$. The following are equivalent:*

(1) *e is extendable;*
(2) *e admits a realisation by a deterministic hidden-variable model;*
(3) *e admits a realisation by a factorisable hidden-variable model.*

---

[11] Note that, due to the assumption of parameter independence (Eq. (6)), we can unambiguously write $k_x(\lambda, -)$ for $k_C|_{\{x\}}(\lambda, -)$, as this marginal is independent of the context C from which one is restricting.

Proof. We prove the sequence of implications $(1) \Rightarrow (2) \Rightarrow (3) \Rightarrow (1)$.

$(1) \Rightarrow (2)$. The idea is that $\mathcal{E}(X) = \boldsymbol{O}_X$ provides a canonical deterministic hidden-variable space. Suppose that e is extendable to a global probability measure μ. Let us set

$$\boldsymbol{\Lambda} := \boldsymbol{O}_X$$
$$p := \mu$$
$$k_C(\mathbf{g}, -) := \delta_{\mathbf{g}|_C}$$

for all global outcome assignments $\mathbf{g} \in \boldsymbol{O}_X$. This is by construction a deterministic hidden-variable model, which we claim gives rise to the empirical model e.

Let $C \in \mathcal{M}$ and write $\rho \colon \boldsymbol{O}_X \longrightarrow \boldsymbol{O}_C$ for the measurable projection which, in the event sheaf, is the restriction map $\rho_C^X = \mathcal{E}(C \subseteq X) \colon \mathcal{E}(X) \longrightarrow \mathcal{E}(C)$.

For any $E \in \mathcal{F}_C$, we have

$$k_C(\mathbf{g}, E) = \delta_{\mathbf{g}|_C}(E) = \delta_{\rho(\mathbf{g})}(E) = \chi_E(\rho(\mathbf{g})) = (\chi_E \circ \rho)(\mathbf{g}) \tag{7}$$

and therefore, as required,

$$\int_{\boldsymbol{\Lambda}} k_C(-, E) \ \mathrm{d}\, p$$
$$= \{ \ \boldsymbol{\Lambda} = \boldsymbol{O}_X; \ p = \mu; \ k_C(-, E) = \chi_E \circ \rho \text{ by Eq. (7)} \ \} \tag{8}$$

$$\int_{\boldsymbol{O}_X} \chi_E \circ \rho \ \mathrm{d}\, \mu$$
$$= \{ \text{ change of variables, Eq. (4) } \} \tag{9}$$

$$\int_{\boldsymbol{O}_C} \chi_E \ \mathrm{d}\, \rho_* \mu$$
$$= \{ \text{ marginalisation for probability measures } \} \tag{10}$$

$$\int_{\boldsymbol{O}_C} \chi_E \ \mathrm{d}\, \mu|_C$$
$$= \{ \text{ integral of indicator function, Eq. (3) } \} \tag{11}$$

$$\mu|_C(E)$$
$$= \{ \ \mu \text{ extends the empirical model } e \ \}$$
$$e_C(E) \ . \tag{12}$$

$(2) \Rightarrow (3)$. It is enough to show that if a hidden-variable model $\langle \boldsymbol{\Lambda}, p, k \rangle$ is deterministic then it is also factorisable. For this, it is sufficient to notice that a Dirac measure $\delta_{\mathbf{o}}$ with $\mathbf{o} \in \boldsymbol{O}_C$ on a product space $\boldsymbol{O}_C = \prod_{x \in C} \boldsymbol{O}_x$ factorises as the product of Dirac measures

$$\delta_{\mathbf{o}} = \prod_{x \in C} \delta_{\mathbf{o}(x)} = \prod_{x \in C} \delta_{\mathbf{o}|_{\{x\}}} \ .$$

$(3) \Rightarrow (1)$. Suppose that e is realised by a factorisable hidden-variable model $\langle \boldsymbol{\Lambda}, p, k \rangle$. Write k_x for $k_C|_{\{x\}}$ as in the definition of factorisability. Define a measure μ on $\boldsymbol{O}_X$ as follows: given a family of measurable sets $(E_x \in \mathcal{F}_x)_{x \in X}$ with $E_x = \boldsymbol{O}_x$ for all but

finitely many $x \in X$, the value of μ on the corresponding cylinder, $\prod_{x \in X} E_x$, is given by

$$\mu\left(\prod_{x \in X} E_x\right) := \int_\Lambda \left(\prod_{x \in X} k_x(-, E_x)\right) \mathrm{d}\, p = \int_{\lambda \in \Lambda} \left(\prod_{x \in X} k_x(\lambda, E_x)\right) \mathrm{d}\, p(\lambda)\,, \quad (13)$$

where the product on the right-hand side is a product of finitely many real numbers in the interval $[0, 1]$, since $k_x(\lambda, O_x) = 1$ and so $k_x(\lambda, E_x) \neq 1$ for only finitely many $x \in X$. Note that the σ-algebra of $\boldsymbol{O}_X$ is the tensor product σ-algebra $\mathcal{F}_X = \bigotimes_{x \in X} \mathcal{F}_x$, which is generated by such cylinders; hence the equation above uniquely determines μ as a measure on $\boldsymbol{O}_X$.

Now, we show that this is a global section for the empirical probabilities. Let $C \in \mathcal{M}$ and consider a 'cylinder' set $F = \prod_{x \in C} F_x$ with $F_x \in \mathcal{F}_X$ and $F_x \neq O_x$ only for finitely many $x \in C$. Then

$$\mu|_C(F)$$
$$= \{ \text{ definition of marginalisation } \} \quad\quad (14)$$
$$\mu(F \times O_{X \setminus C})$$
$$= \{ \text{ definition of } F \text{ and } O_U \} \quad\quad (15)$$
$$\mu(\prod_{x \in C} F_x \times \prod_{x \in X \setminus C} O_x)$$
$$= \{ \text{ definition of } \mu, \text{ Eq. (13) } \} \quad\quad (16)$$
$$\int_\Lambda \left(\prod_{x \in C} k_C(-, F_x)\right) \left(\prod_{x \in X \setminus C} k_C(-, O_x)\right) \mathrm{d}\, p$$
$$= \{ k_C(\lambda, -) \text{ is a probability measure so } k_C(\lambda, O_x) = 1 \} \quad\quad (17)$$
$$\int_\Lambda \left(\prod_{x \in C} k_C(-, F_x)\right) \mathrm{d}\, p$$
$$= \{ \text{ factorisability of the hidden-variable model } \} \quad\quad (18)$$
$$\int_\Lambda k_C(-, \prod_{x \in C} F_x) \mathrm{d}\, p$$
$$= \{ \text{ definition of } F \} \quad\quad (19)$$
$$\int_\Lambda k_C(-, F) \mathrm{d}\, p$$
$$= \{ e \text{ is the empirical model corresponding to } \langle \Lambda, p, k \rangle \}$$
$$e_C(F) \quad\quad (20)$$

Since the σ-algebra $\mathcal{F}_C$ of $\boldsymbol{O}_C$ is generated by the cylinder sets of the form above and we have seen that $\mu|_C$ agrees with e_C on these sets, we conclude that $\mu|_C = e_C$ as required. $\quad\square$

5. Quantifying Contextuality

Beyond questioning whether a given empirical behaviour is contextual or not, it is also interesting to ask to what *degree* it is contextual. In discrete-variable scenarios, a very natural measure of contextuality is the contextual fraction [8]. This measure was shown in [6] to have a number of very desirable properties. It can be calculated using linear programming, an approach that subsumes the more traditional approach to quantifying nonlocality and contextuality using Bell and noncontextuality inequalities in the sense that we can understand the (dual) linear program as optimising over *all* such inequalities for the scenario in question and returning the maximum normalised violation of *any* Bell or noncontextuality inequality achieved by the given empirical model. Crucially, the contextual fraction was also shown to *quantifiably* relate to quantum-over-classical advantages in specific informatic tasks [6,63,85]. Moreover it has been demonstrated to be a monotone with respect to the free operations of resource theories for contextuality [4,6,32].

In this section, we consider how to carry those ideas to the continuous-variable setting. The formulation as a linear optimisation problem and the attendant correspondence with Bell inequality violations requires special care as one needs to use infinite linear programming, necessitating some extra assumptions on the outcome measurable spaces.

5.1. The contextual fraction

Asking whether a given behaviour is noncontextual amounts to asking whether the empirical model is extendable, or in other words whether it admits a deterministic hidden-variable model. However, a more refined question to pose is *what fraction of the behaviour admits a deterministic hidden-variable model?* This quantity is what we call the noncontextual fraction. Similarly, the fraction of the behaviour that is left over and that can thus be considered *irreducibly* contextual is what we call the contextual fraction.

Definition 8. Let e be an empirical model on the scenario $\langle X, \mathcal{M}, \boldsymbol{O} \rangle$. The *noncontextual fraction* of e, written $\mathsf{NCF}(e)$, is defined as

$$\sup \{\mu(O_X) \mid \mu \in \mathbb{M}(\boldsymbol{O}_X), \forall C \in \mathcal{M}. \; \mu|_C \leq e_C\} \; .$$

Note that since $e_C \in \mathbb{P}(\boldsymbol{O}_C)$ for all $C \in \mathcal{M}$ it follows that $\mathsf{NCF}(e) \in [0, 1]$. The *contextual fraction* of e, written $\mathsf{CF}(e)$, is given by $\mathsf{CF}(e) := 1 - \mathsf{NCF}(e)$.

5.2. Monotonicity under free operations including binning

In the discrete-variable setting, the contextual fraction was shown to be a monotone under a number of natural classical operations that transform and combine empirical models and control their use as resources, therefore constituting the 'free' operations of a resource theory of contextuality [4,6,32].

All of the operations defined for discrete variables in [6]—viz. translations of measurements, transformation of outcomes, probabilistic mixing, product, and choice—carry almost verbatim to our current setting. One detail is that one must insist that the coarse-graining of outcomes be achieved by (a family of) measurable functions. A particular example of practical importance is *binning*, which is widely used in continuous-variable quantum information as a method of discretising data by partitioning the outcome space $\boldsymbol{O}_x$ for each measurement $x \in X$ into a finite number of 'bins', i.e. mea-

surable sets. Note that a binned empirical model is obtained by pushing forward along a family $(t_x)_{x \in X}$ of outcome translations $t_x : \boldsymbol{O}_x \longrightarrow \boldsymbol{O}'_x$ where $\boldsymbol{O}'_x$ is finite for all $x \in X$.

For the conditional measurement operation introduced in [4], which allows for adaptive measurement protocols such as those used in measurement-based quantum computation [76], one must similarly insist that the map determining the next measurement to perform based on the observed outcome of a previous measurement be a measurable function. Since, for the quantification of contextuality, we are only considering scenarios where the measurements are treated as constituting a discrete set, this amounts to a partition of the outcome space $\boldsymbol{O}_x$ of the first measurement, x, into measurable subsets labelled by measurements compatible with x, indicating which will be subsequently performed depending on the outcome observed for x.

The inequalities establishing monotonicity from [6, Theorem 2] also hold for continuous variables. There is a caveat for the equality formula for the product of two empirical models:

$$\mathsf{NCF}(e_1 \otimes e_2) = \mathsf{NCF}(e_1)\mathsf{NCF}(e_2).$$

Whereas the inequality establishing monotonicity ($\geq$) stills holds in general, the proof establishing the other direction ($\leq$) makes use of duality of linear programs. Therefore, it only holds under the assumptions we will impose in the remainder of this section.

Proposition 1. *If e is an empirical model, and e^{bin} is any discrete-variable empirical model obtained from e by binning, then contextuality of e^{bin} witnesses contextuality of e, and quantifiably gives a lower bound $\mathsf{CF}(e^{bin}) \leq \mathsf{CF}(e)$.*

5.3. Assumptions on the outcome spaces

In order to phrase the problem of contextuality as an (infinite) linear programming problem and establish the connection with violations of Bell inequalities, we need to impose some conditions on the measurement scenarios, and in particular on the measurable spaces of outcomes.

First, from now on we assume that we have a *finite number of measurement labels* i.e. that X is finite.

Moreover, we restrict attention to the case where the outcome space $\boldsymbol{O}_x$ for each measurement $x \in X$ is the Borel measurable space for a compact Hausdorff space, i.e. that the set O_x is a compact space and $\mathcal{F}_x$ is the σ-algebra generated by its open sets, written $\mathcal{B}(O_x)$. Note that this includes most situations of interest in practice. In particular, it includes the case of measurements with outcomes in a bounded subspace of $\mathbb{R}$ or $\mathbb{R}^n$. This is also experimentally motivated since measurement devices are energetically bounded. The central missing piece is the case of locally compact spaces, in order to include the measurements with outcomes in $\mathbb{R}$ or $\mathbb{R}^n$, which is theoretically relevant ($\mathbb{R}$ would be the canonical outcome space for the quadratures of the electromagnetic field, for instance). We address this issue in the next section and show that it reduces to the compact case.

To summarise we make the following two assumptions here (we will slightly relax the second one later):

(i) X is a finite set of measurement labels,
(ii) for each $x \in X$, the outcome space $\boldsymbol{O}_x$ is a compact Hausdorff space.

To obtain an infinite linear program, we need to work with vector spaces. However, probability measures, or even finite or arbitrary measures, do not form one. We will therefore consider the set $\mathbb{M}_{\pm}(Y)$ of *finite signed measures* (a.k.a. real measures) on a measurable space $Y = \langle Y, \mathcal{F}_Y \rangle$. These are functions $\mu \colon \mathcal{F}_Y \longrightarrow \mathbb{R}$ such that $\mu(\emptyset) = 0$ and μ is σ-additive. In comparison to the definition of a measure, one drops the nonnegativity requirement, but insists that the values be finite. The set $\mathbb{M}_{\pm}(Y)$ forms a real vector space which includes the probability measures $\mathbb{P}(Y)$, and total variation gives a norm on this space. When Y is a compact Hausdorff space and $Y = \langle Y, \mathcal{B}(Y) \rangle$, the Riesz–Markov–Kakutani representation theorem [48] says that $\mathbb{M}_{\pm}(Y)$ is a concrete realisation of the topological dual space of $C(Y, \mathbb{R})$, the space of continuous real-valued functions on Y. The duality is given by $\langle \mu, f \rangle := \int_Y f \ \mathrm{d}\mu$ for $\mu \in \mathbb{M}_{\pm}(Y)$ and $f \in C(Y, \mathbb{R})$.

5.4. Linear programming

Consider an empirical model $e = (e_C)_{C \in \mathcal{M}}$ on a scenario $\langle X, \mathcal{M}, O \rangle$ satisfying the assumptions discussed above. Calculation of its noncontextual fraction can be expressed as the infinite linear programming problem (P-CF). This is our primal linear program; its dual linear program is given by (D-CF). In what follows, we will see how to derive the dual and show that the optimal solutions of both programs coincide. We also refer the interested reader to Appendix A where the programs are expressed in the standard form of infinite linear programming [17].

$$(\text{P-CF}) \quad \begin{cases} \text{Find } \mu \in \mathbb{M}_{\pm}(O_X) \\ \text{maximising } \mu(O_X) \\ \text{subject to:} \\ \forall C \in \mathcal{M}, \ \mu|_C \ \leq \ e_C \\ \mu \ \geq \ 0 \, . \end{cases}$$

$$(\text{D-CF}) \quad \begin{cases} \text{Find } (f_C)_{C \in \mathcal{M}} \in \prod_{C \in \mathcal{M}} C(O_C) \\ \text{minimising } \sum_{C \in \mathcal{M}} \int_{O_C} f_C \ \mathrm{d}\, e_C \\ \text{subject to:} \\ \sum_{C \in \mathcal{M}} f_C \circ \rho_C^X \ \geq \ \mathbf{1}_{O_X} \\ \forall C \in \mathcal{M}, \ f_C \ \geq \ \mathbf{0}_{O_C} \, . \end{cases}$$

We have written ρ_C^X for the projection $O_X \longrightarrow O_C$ as before, and $\mathbf{1}_D$ (resp. $\mathbf{0}_D$) for the constant function $D \longrightarrow \mathbb{R}$ that assigns the number 1 (resp. 0) to all elements of its domain D; in the above instance, to all $\mathbf{g} \in O_X$ (resp. all $\mathbf{o} \in O_C$).[12] We denote the optimal values of problems (P-CF) and (D-CF), respectively, as val(P-CF) and val(D-CF). They both equal $\mathsf{NCF}(e)$ due to strong duality (see Proposition 24 and Appendix B).

---

[12] Note that $\mathbf{1}_D$ is just a simplified notation for the indicator function on the whole domain; i.e. $\mathbf{1}_D = \chi_D \colon D \longrightarrow \mathbb{R}$. Similarly, $\mathbf{0}_D$ is the indicator function of the empty set; i.e. $\mathbf{0}_D = \chi_\emptyset \colon D \longrightarrow \mathbb{R}$.

Analogues of these programs have been studied in the discrete-variable setting [6]. Note however that, in general, these continuous-variable linear programs are over infinite-dimensional spaces and thus not practical to compute directly. For this reason, in Sect. 8 we will introduce a hierarchy of finite-dimensional semi-definite programs that approximate the solution of (P-CF) to arbitrary precision.

Deriving the dual via the Lagrangian

We now give an explicit derivation of (D-CF) as the dual of (P-CF) via the Lagrangian method. To simplify notation, we set $E_1 := \mathbb{M}_{\pm}(\boldsymbol{O}_X)$ and $F_2 := \prod_{C \in \mathcal{M}} C(O_C, \mathbb{R})$ and their convex cones K_1 and K_2^* (see Appendix A). This matches the standard form notation for infinite linear programming of [17], in which we present our programs in Appendix A. Hence we introduce $|\mathcal{M}|$ dual variables, and one continuous map $f_C \in C(O_C, \mathbb{R})$ for each $C \in \mathcal{M}$ to account for the constraints $\mu|_C \le e_C$. From (P-CF), we then define the Lagrangian $\mathcal{L} : K_1 \times K_2^* \longrightarrow \mathbb{R}$ as

$$\mathcal{L}(\mu, (f_C)) := \underbrace{\mu(O_X)}_{\text{objective}} + \underbrace{\sum_{C \in \mathcal{M}} \int_{O_C} f_C \; \mathrm{d}(e_C - \mu|_C)}_{\text{constraints}} . \tag{21}$$

The primal program (P-CF) corresponds to

$$\sup_{\mu \in K_1} \inf_{(f_C) \in K_2^*} \mathcal{L}(\mu, (f_C)) , \tag{22}$$

as the infimum here imposes the constraints that $\mu|_C \le e_C$ for all $C \in \mathcal{M}$, for otherwise the Lagrangian diverges. If these constraints are satisfied, then because of the infimum, the second term of the Lagrangian vanishes yielding the objective of the primal problem. To express the dual, which amounts to permuting the infimum and the supremum, we need to rewrite the Lagrangian:

$$
\begin{aligned}
\mathcal{L}(\mu, (f_C)) &= \mu(O_X) + \sum_{C \in \mathcal{M}} \int_{O_C} f_C \; \mathrm{d}(e_C - \mu|_C) \\
&= \int_{O_X} \mathbf{1} \; \mathrm{d}\mu + \sum_{C \in \mathcal{M}} \int_{O_C} f_C \; \mathrm{d}e_C - \sum_{C \in \mathcal{M}} \int_{O_C} f_C \; \mathrm{d}\mu|_C \\
&= \int_{O_X} \mathbf{1} \; \mathrm{d}\mu + \sum_{C \in \mathcal{M}} \int_{O_C} f_C \; \mathrm{d}e_C - \sum_{C \in \mathcal{M}} \int_{O_X} f_C \circ \rho_C^X \; \mathrm{d}\mu \\
&= \int_{O_X} \mathbf{1} \; \mathrm{d}\mu + \sum_{C \in \mathcal{M}} \int_{O_C} f_C \; \mathrm{d}e_C - \int_{O_X} \left(\sum_{C \in \mathcal{M}} f_C \circ \rho_C^X \right) \mathrm{d}\mu \\
&= \sum_{C \in \mathcal{M}} \int_{O_C} f_C \; \mathrm{d}e_C + \int_{O_X} \left(\mathbf{1} - \sum_{C \in \mathcal{M}} f_C \circ \rho_C^X \right) \mathrm{d}\mu .
\end{aligned}
$$

The dual program (D-CF) indeed corresponds to

$$\inf_{(f_C) \in K_2^*} \sup_{\mu \in K_1} \mathcal{L}(\mu, (f_C)) . \tag{23}$$

The supremum imposes that $\sum_{C \in \mathcal{M}} f_C \circ \rho_C^X \geq \mathbf{1}$ on O_X, since otherwise the Lagrangian diverges. If this constraint is satisfied, then the supremum makes the second term vanish yielding the objective of the dual problem (D-CF).

Zero duality gap

A key result about the noncontextual fraction, which is essential in establishing the connection to Bell inequality violations, is that (P-CF) and (D-CF) are strongly dual, in the sense that no gap exists between their optimal values. Strong duality always holds in finite linear programming, but it does not hold in general for the infinite case.

Proposition 2. *Problems* (P-CF) *and* (D-CF) *have zero duality gap and their optimal values satisfy*

$$\text{val(P-CF)} = \text{val(D-CF)} = \mathsf{NCF}(e) \tag{24}$$

Proof. This proof relies on [17, Theorem 7.2]. The complete proof is provided in Appendix B. Here, we only provide a brief outline. Let $E_1 := \mathbb{M}_{\pm}(\boldsymbol{O}_X) \times \prod_{C \in \mathcal{M}} \mathbb{M}_{\pm}(\boldsymbol{O}_C)$ and $E_2 := \prod_{C \in \mathcal{M}} \mathbb{M}_{\pm}(\boldsymbol{O}_C)$. Strong duality between (P-CF) and (D-CF) amounts to showing that the cone

$$\mathcal{K} = \{ ((\mu|_C + \nu_C)_{C \in \mathcal{M}}, \mu(O_X)) \mid (\mu, (\nu_C)_{C \in \mathcal{M}}) \in E_{1+} \}$$

is weakly closed in $E_2 \oplus \mathbb{R}$, where:

$$E_{1+} := \{ (\mu, (\nu_C)_{C \in \mathcal{M}}) \in E_1 \mid \mu \geq 0 \text{ and } \forall C \in \mathcal{M}.\ \nu_C \geq 0 \} \subset E_1.$$

We do so by considering a sequence $(\mu^k, (\nu_C^k)_C)_{k \in \mathbb{N}}$ in E_{1+} and showing that the accumulation point

$$\lim_{k \to \infty} \left((\mu^k|_C + \nu^k)_{C \in \mathcal{M}}, \mu^k(O_X) \right)$$

belongs to $\mathcal{K}$. $\square$

6. The Case of Local Compactness

We now focus on cases where the outcome space might be only locally compact. These include most theoretical situations that are of interest in practice. For instance $\mathbb{R}$ could be the outcome space for the position and momentum operators.

For each measurement $x \in X$, $\boldsymbol{O}_x$ is supposed to be the Borel measurable space for a second-countable locally compact Hausdorff space, i.e. that the set O_x is equipped with a second-countable locally compact Hausdorff topology and $\mathcal{F}_x$ is the σ-algebra generated by its open sets, written $\mathcal{B}(O_x)$. Second countability and Hausdorffness of two spaces Y and Z suffice to show that the Borel σ-algebra of the product topology is the tensor product of the Borel σ-algebras, i.e. $\mathcal{B}(Y \times Z) = \mathcal{B}(Y) \otimes \mathcal{B}(Z)$ [22, Lemma 6.4.2 (Vol. 2)]. Hence, these assumptions guarantee that $\boldsymbol{O}_U$ for $U \in \mathcal{P}(X)$ is the Borel σ-algebra of the product topology on $O_U = \prod_{x \in U} O_x$. These product spaces are also second-countable, locally compact, and Hausdorff as all three properties are preserved by finite products. When Y is a second-countable locally compact Hausdorff space and $\boldsymbol{Y} = \langle Y, \mathcal{B}(Y) \rangle$, the Riesz–Markov–Kakutani representation theorem [48]

says that $\mathbb{M}_{\pm}(Y)$ is a concrete realisation of the topological dual space of $C_0(Y)$, the space of continuous real-valued functions on Y that vanish at infinity.[13] The duality is given by $\langle \mu, f \rangle := \int_Y f \, \mathrm{d}\mu$ for $\mu \in \mathbb{M}_{\pm}(Y)$ and $f \in C_0(Y)$.[14] Note that when Y is compact (as treated above), $C_0(Y) = C(Y)$ as every closed subspace of a compact space is compact.

Next, we show that we can approximate the linear program (P-CF)[15] by a slightly modified linear program defined on the space of finite measures on a measurable compact subspace of O_X. The idea is to approximate to any desired error the mass of a finite measure on a locally compact set by the mass of the same measure on a compact subset. This naturally comes from the notion of *tightness* of a measure.

Definition 9 (tightness of a measure). A measure μ on a metric space U is said to be *tight* if for each $\varepsilon > 0$ there exists a compact set $U_\varepsilon \subseteq U$ such that $\mu(U \setminus U_\varepsilon) < \varepsilon$.

Then we need to argue that every measure we will consider is tight. This is a result of the following theorem.

Theorem 2 [70]. *If S is a complete separable metric space, then every finite measure on S is tight.*

For $x \in X$, O_x is a second-countable locally compact Hausdorff space, thus a Polish space i.e. a separable completely metrisable topological space. For this reason, the above theorem applies. We are now ready to state and prove the main theorem of this subsection.

Theorem 3. *The linear program* (PC-F$^{\mathrm{CV},\varepsilon}$)*, defined over finite-signed measures on a locally compact space can be approximated to any desired precision ε by a linear program* (P-CF$^{\mathrm{CV},\varepsilon}$) *defined over finite signed measures on a compact space.*

Proof. Fix $\varepsilon > 0$. Let $C \in \mathcal{M}$ be a given context and $x \in C$ a given measurement label within that context. Because e_C is a probability measure on O_C, the marginal measure $e_C|_{\{x\}}$ is a finite measure on O_x. Following Theorem 2, $e_C|_{\{x\}}$ is tight and there exists a compact subset $K_x^{\varepsilon,C} \subseteq O_x$ such that: $e_C|_{\{x\}}(O_x \setminus K_x^{\varepsilon,C}) \leq \varepsilon$. Importantly there exist proofs that explicitly construct the approximating sets $K_x^{\varepsilon,C}$ (see [67]) based on the separability of the underlying spaces. It makes this construction feasible in practice and justifies this approach.

We apply this procedure for every context and for all measurements in a context. We now define the compact set

$$O_x^\varepsilon := \bigcup_{C|x\in C} K_x^{\varepsilon,C}.$$

The previous definition is essential to ensure a noncontextual cut off of the outcome set which ensures the good definition of a compact subset for each measurement label independent of the context. For some subset of measurement labels $U \subseteq X$, we define

---

[13] A function $f \colon Y \longrightarrow \mathbb{R}$ on a locally compact space Y is said to *vanish at infinity* if the set $\{y \in Y \mid \|f(x)\| \geq \varepsilon\}$ is compact for all $\varepsilon > 0$.

[14] This theorem holds more generally for locally compact Hausdorff spaces if one considers only (finite signed) Radon measures, which are measures that play well with the underlying topology. However, second-countability, together with local compactness and Hausdorffness, guarantees that every Borel measure is Radon [37, Theorem 7.8].

[15] Here we will still use the form of the program (P-CF) though throughout this subsection one has to keep in mind that it is defined over finite-signed measures on a *locally compact* space rather than a compact space.

the compact set $O_U^\varepsilon := \prod_{x \in U} O_x^\varepsilon$. For every context $C \in \mathcal{M}$ and for every measurement label $x \in C$, we now have that $K_x^{\varepsilon, C} \subseteq O_x^\varepsilon$ and thus $e_C|_{\{x\}}(O_x \setminus O_x^\varepsilon) \le \varepsilon$. Note that due to the compatibility condition, we can write $e_C|_{\{x\}}$ as $e_{\{x\}}$ for any context.

Let μ be any feasible solution of (P-CF) defined over finite-signed measures on a locally compact space. Due to the constraints of (P-CF) we have that $\forall x \in X$, $\mu|_{\{x\}} \le e_{\{x\}}$. Then:

$$\mu(O_X \setminus O_X^\varepsilon) = \mu \left(\prod_{x \in X} O_x \setminus \prod_{x \in X} O_x^\varepsilon \right) \tag{25}$$

$$= \mu \left(\prod_{x \in X} (O_x \setminus O_x^\varepsilon) \right) \tag{26}$$

$$= \prod_{x \in X} \mu|_{\{x\}}(O_x \setminus O_x^\varepsilon) \tag{27}$$

$$\le \prod_{x \in X} e_{\{x\}}(O_x \setminus O_x^\varepsilon) \tag{28}$$

$$\le \varepsilon^{|X|}. \tag{29}$$

We now define the linear program (P-CF$^{\mathrm{CV}, \varepsilon}$) which has the same form as (P-CF) though the unknown measures are taken from $\mathbb{M}_\pm(\boldsymbol{O}_X^\varepsilon)$ where $\boldsymbol{O}_X^\varepsilon = \langle O_X^\varepsilon, \mathcal{B}(O_X^\varepsilon) \rangle$. We would like to state that (P-CF$^{\mathrm{CV}, \varepsilon}$) approximates (P-CF) up to ε; i.e. that their values are ε-close. The missing ingredient from the previous chain of inequalities is that given an optimal measure μ^* satisfying (P-CF), we do not know whether an optimal solution μ_ε^* of (P-CF$^{\mathrm{CV}, \varepsilon}$) is necessarily the restriction of μ^* to O_X^ε. In fact, it is possible that we do not even have a unique optimal solution. However we only need to prove that they have the same mass on O_X^ε, i.e. $\mu_\varepsilon^*(O_X^\varepsilon) = \mu^*|_{O_X^\varepsilon}(O_X^\varepsilon)$. For a contradiction, suppose this does not hold. Then because μ_ε^* is an optimal value of (P-CF$^{\mathrm{CV}, \varepsilon}$), we must have $\mu_\varepsilon^*(O_X^\varepsilon) > \mu^*|_{O_X^\varepsilon}(O_X^\varepsilon)$. From this we construct a new measure $\tilde{\mu}$ on $\boldsymbol{O}_X$ which equals μ_ε^* on $\boldsymbol{O}_X^\varepsilon$ and μ^* on $\boldsymbol{O}_X \setminus O_X^\varepsilon$. It satisfies all constraints and furthermore $\tilde{\mu}(O_X) > \mu^*(O_X)$. This contradicts the fact that μ^* is an optimal solution of (P-CF). Thus necessarily $\mu_\varepsilon^*(O_X^\varepsilon) = \mu^*|_{O_X^\varepsilon}(O_X^\varepsilon)$.

The linear program (P-CF$^{\mathrm{CV}, \varepsilon}$) defined on a compact space has indeed a value ε-close to the original program (P-CF). $\square$

In conclusion to this section, we can approximate the problem of finding the noncontextual fraction in measurement scenarios whose outcome spaces are locally compact by the same problem defined on compact subspace. It thus suffices to restrict the study to the case of compact outcome spaces.

7. Continuous Generalisation of Bell Inequalities

The dual program (D-CF) is of particular interest in its own right. As we now show, it can essentially be understood as computing a continuous-variable 'Bell inequality' that is optimised to the empirical model. Making the change of variables $\beta_C := |\mathcal{M}|^{-1} \mathbf{1}_{O_C} - f_C$

for each $C \in \mathcal{M}$, the dual program (D-CF) transforms to the following.

$$(\text{B-CF}) \begin{cases} \text{Find } (\beta_C)_{C \in \mathcal{M}} \in \prod_{C \in \mathcal{M}} C(O_C) \\[2ex] \text{maximising } \sum_{C \in \mathcal{M}} \int_{O_C} \beta_C \, d\,e_C \\[2ex] \text{subject to:} \\[1ex] \sum_{C \in \mathcal{M}} \beta_C \circ \rho_C^X \leq \mathbf{0}_{O_X} \\[2ex] \forall C \in \mathcal{M}, \ \beta_C \leq |\mathcal{M}|^{-1} \mathbf{1}_{O_C} \, . \end{cases}$$

This program directly computes the contextual fraction $\mathsf{CF}(e)$ instead of the noncontextual fraction. It maximises, subject to constraints, the total value obtained by integrating these functionals context-wise against the empirical model in question. The first set of constraints—a generalisation of a system of linear inequalities determining a Bell inequality — ensures that, for noncontextual empirical models, the value of the program is at most 0, since any such model extends to a measure μ on O_X such that $\mu(O_X) = 1$. The final set of constraints acts as a normalisation condition on the value of the program, ensuring that it takes values in the interval $[0, 1]$ for any empirical model. Any family of functions $\beta = (\beta_C) \in F_2$ satisfying the constraints will thus result in what can be regarded as a generalised Bell inequality,

$$\sum_{C \in \mathcal{M}} \int_{O_C} \beta_C \, d\,e_C \leq 0, \tag{30}$$

which is satisfied by all noncontextual empirical models.

Definition 10. A *form* β on a measurement scenario $\langle X, \mathcal{M}, \mathbf{O} \rangle$ is a family $\beta = (\beta_C)_{C \in \mathcal{M}}$ of functions $\beta_C \in C(O_C)$ for all $C \in \mathcal{M}$. Given an empirical model e on $\langle X, \mathcal{M}, \mathbf{O} \rangle$, the *value* of β on e is[16]

$$\langle \beta, e \rangle_2 := \sum_{C \in \mathcal{M}} \int_{O_C} \beta_C \, d\,e_C \, .$$

The norm of β is given by

$$\|\beta\| := \sum_{C \in \mathcal{M}} \|\beta_C\| = \sum_{C \in \mathcal{M}} \sup \{\beta_C(\mathbf{o}) \mid \mathbf{o} \in O_C\} \, .$$

Definition 11. An *inequality* (β, R) on a measurement scenario $\langle X, \mathcal{M}, \mathbf{O} \rangle$ is a form β together with a bound $R \in \mathbb{R}$. An empirical model e is said to satisfy the inequality if the value of β on e is below the bound, i.e. $\langle \beta, e \rangle_2 \leq R$.

Definition 12. An inequality (β, R) is said to be a *generalised Bell inequality* if it is satisfied by all noncontextual empirical models, i.e. if for any noncontextual model d on $\langle X, \mathcal{M}, \mathbf{O} \rangle$, it holds that $\langle \beta, d \rangle_2 \leq R$.

---

[16] The notation $\langle \cdot, \cdot \rangle_2$ is further discussed and explained to be a canonical duality in Appendix A.

A generalised Bell inequality (β, R) establishes a bound $\langle e, \beta \rangle_2$ amongst noncontextual models e. For more general models, the value of β on e is only limited by the algebraic bound $\|\beta\|$. In the following, we will only consider inequalities (β, R) for which $R < \|\beta\|$ excluding inequalities trivially satisfied by all empirical models.

Definition 13. The *normalised violation* of a generalised Bell inequality (β, R) by an empirical model e is

$$\frac{\max\{0, \langle \beta, e \rangle_2 - R\}}{\|\beta\| - R},$$

the amount by which its value $\langle \beta, e \rangle_2$ exceeds the bound R normalised by the maximal 'algebraic' violation.

The above definition restricts to the usual notions of Bell inequality and noncontextual inequality in the discrete-variable case and is particularly close to the presentation in [6]. The following theorem also generalises to continuous variables the main result of [6].

Theorem 4. *Let e be an empirical model. (i) The normalised violation by e of any generalised Bell inequality is at most $\mathsf{CF}(e)$; (ii) if $\mathsf{CF}(e) > 0$ then for every $\varepsilon > 0$ there exists a generalised Bell inequality whose normalised violation by e is at least $CF(e) - \varepsilon$.*

Proof. The proof follows directly from the definitions of the linear programs, and from strong duality, i.e. the fact that their optimal values coincide (Proposition 2 below). □

Item (ii) is slightly modified compared to the discrete analogue because there is no guarantee that there exists an optimal solution for the dual program (D-CF). In particular, its optimal value might be achieved by a discontinuous function that can be approximated by continuous ones. Hence the modification of (ii) with a normalised violation ε-close to $\mathsf{CF}(e)$.

8. Approximating the Contextual Fraction with SDPs

In Sect. 5, we presented the problem of computing the noncontextual fraction as an infinite linear program. Although this is of theoretical importance, it does not allow one to directly perform the actual numerical computation of this quantity. Here we exploit the link between measures and their sequence of moments to derive a hierarchy of truncated finite-dimensional semidefinite programs which are a relaxation of the original primal problem (P-CF). Dual to this vision, we can equivalently exploit the link between positive polynomials and their sum-of-squares representation to derive a hierarchy of semidefinite programs which are a restriction of the dual problem (D-CF). We further prove that the optimal values of the truncated programs converge monotonically to the noncontextual fraction. This makes use of global optimisation techniques developed by Lasserre and Parrilo [54,69] and further developed in [55]. We introduce them in Appendix C and strongly recommend reading this appendix to readers unfamiliar with these notions. We will use the same notation throughout this section. Another extensive and well-presented reference on the subject is [56]. We start by deriving a hierarchy of SDPs to approximate the contextual fraction and then show that it provides a sequence of optimal values that converge to the noncontextual fraction.

Notation and terminology

We first fix some notation that is also used in Appendix C. Let $\mathbb{R}[\boldsymbol{x}]$ denote the ring of real polynomials in the variables $\boldsymbol{x} \in \mathbb{R}^d$, and let $\mathbb{R}[\boldsymbol{x}]_k \subset \mathbb{R}[\boldsymbol{x}]$ contain those polynomials of total degree at most k. The latter forms a vector space of dimension $s(k) := \binom{d+k}{k}$, with a canonical basis consisting of monomials $\boldsymbol{x}^{\boldsymbol{\alpha}} = x_1^{\alpha_1} \cdots x_d^{\alpha_d}$ indexed by the set $\mathbb{N}_k^d := \left\{ \boldsymbol{\alpha} \in \mathbb{N}^d \mid |\boldsymbol{\alpha}| \leq k \right\}$ where $|\boldsymbol{\alpha}| := \sum_{i=1}^d \alpha_i$. Any $p \in \mathbb{R}[\boldsymbol{x}]_k$ can be expanded in this basis as $p(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha} \in \mathbb{N}_k^d} p_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}}$ and we write $\mathbf{p} := (p_{\boldsymbol{\alpha}}) \in \mathbb{R}^{s(k)}$ for the resulting vector of coefficients.

8.1. Hierarchy of semidefinite relaxations for computing $\mathsf{NCF}(e)$

We fix a measurement scenario $\langle X, \mathcal{M}, \boldsymbol{O} \rangle$ and an empirical model e on this scenario. We will restrict our attention to outcome spaces of the form detailed in Sect. 5.3. Let $d = |X| \in \mathbb{N}_{>0}$ so that O_X is a Borel subset of $\mathbb{R}^d$. As a prerequisite, we first need to compute the sequences of moments associated to measures $(e_C)_{C \in \mathcal{M}}$ derived from the empirical model. For $C \in \mathcal{M}$, let $\boldsymbol{y}^{e,C} = (y_{\boldsymbol{\alpha}}^{e,C})_{\boldsymbol{\alpha} \in \mathbb{N}^d}$ be the sequence of all moments of e_C. For a given $k \in \mathbb{N}$, which will fix the level of the hierarchy, we only need to compute a finite number $s(k)$ of moments for all contexts. These will be the inputs to the program.

Below, we derive a hierarchy of SDP relaxations for the primal program (P-CF) such that their optimal values converge monotonically to val(P-CF) = $\mathsf{NCF}(e)$. We start by discussing the assumptions we have to make on the outcome space. Then we derive the hierarchy based first on the primal program and then on the dual program and we further show that these formulations are indeed dual. Finally, we prove convergence of the hierarchy.

Further assumptions on the outcome space?

We already made the assumptions mentioned in Sect. 5.3 for the outcome spaces $\boldsymbol{O} = (\boldsymbol{O}_x)_{x \in X}$ noting that they are not restrictive when considering actual applications. However we would like to meet the assumptions detailed in Assumption 1 for the global outcome space O_X so that both Theorems 6 and 8 apply in our setting (see Appendix C).

Assumption 1 (ii) is already met because we have assumed that for all $x \in X$, $O_x \subset \mathbb{R}$ is compact. Recall that the more general case of locally compact can be reduced to the compact case, as seen in Sect. 6.

Let us discuss Assumption 1 (i). We have that $O_X = \prod_{x \in X} O_x$ with $O_x \subset \mathbb{R}$ compact. If O_x is disconnected, we can always complete it into a connected space by attributing measure zero to the added parts for all measures e_C whenever $x \in C$. Then because O_x is compact, it is bounded and it can be described by two constant polynomials: there exists $a_x, b_x \in \mathbb{R}$ such that $O_x = [a_x, b_x]$. This makes O_X a polytope so in particular, it is semi-algebraic. We write it as

$$O_X = \left\{ \boldsymbol{x} \in \mathbb{R}^d \mid \forall j = 1, \ldots, m, \; g_j(\boldsymbol{x}) \geq 0 \right\} \tag{31}$$

for some polynomials $g_j \in \mathbb{R}[\boldsymbol{x}]$ of degree 1.

As noted in [54], Assumption 1 (iii) is not very restrictive. For instance, it is satisfied when the set is a polytope. This is the case for O_X.

Thus there is no need for further assumptions beyond those already assumed in Sect. 5.3 in order to apply the results presented in Appendix C.

Relaxation of the primal program

The program (P-CF) can be relaxed so that a converging hierarchy of SDPs can be derived. The program (P-CF) is essentially a maximisation problem on finite-signed Borel measures with additional constraints such as the fact that these are proper measures (i.e. they are nonnegative). We will represent a measure by its moment sequence and use conditions for which this moment sequence has a (unique) representing Borel measure (see Appendix C.2). We recall the expression of the primal program (P-CF):

$$
\text{(P-CF)} \begin{cases}
\text{Find } \mu \in \mathbb{M}_{\pm}(\boldsymbol{O}_X) \\
\text{maximising } \mu(O_X) \\
\text{subject to:} \\
\quad \forall C \in \mathcal{M}, \ \mu|_C \ \leq \ e_C \\
\quad \mu \ \geq \ 0 \, .
\end{cases}
$$

From Appendix C.2 which culminates at Theorem 8, it can be relaxed for $k \in \mathbb{N}_{>0}$ as:

$$
\text{(SDP-CF}^k\text{)} \begin{cases}
\text{Find } \boldsymbol{y} \in \mathbb{R}^{s(2k)} \\
\text{maximising } y_{\boldsymbol{0}} \\
\text{subject to:} \\
\quad \forall C \in \mathcal{M}, \ M_k(\boldsymbol{y}^{e,C} - \boldsymbol{y}|_C) \succeq 0, \\
\quad \forall j = 1, \ldots, m, \ M_{k-1}(g_j \boldsymbol{y}) \succeq 0, \\
\quad M_k(\boldsymbol{y}) \succeq 0 \, .
\end{cases}
$$

The moment matrices $M_k(\boldsymbol{y})$ and the localising matrices $M_{k-1}(g_j\boldsymbol{y})$ are defined in Appendix C. We consider localising matrices of order $k-1$ rather than k because all g_j's are of degree exactly 1. In this way, the maximum degree matches with that of the moment matrices. In general we have to deal with localising matrices of order $k-\lceil \frac{\deg(g_j)}{2} \rceil$. If μ is a representing measure on $\boldsymbol{O}_X$ for $\boldsymbol{y}$ then for all contexts $C \in \mathcal{M}, \boldsymbol{y}|_C$ can be defined through $\boldsymbol{y}$ by requiring that $\boldsymbol{y}|_C$ has representing measure $\mu|_C$. The two last constraints state necessary conditions on the variable $\boldsymbol{y}$ to be moments of some finite Borel measure supported on $\boldsymbol{O}_X$. The first constraint is a relaxation of the constraint $\mu|_C \leq e_C$ for $C \in \mathcal{M}$. As expected, (SDP-CF^k) is a semidefinite relaxation of the problem (P-CF) so that $\forall k \in \mathbb{N}_{>0}$, $\mathsf{NCF}(e) = \text{val(P-CF)} \leq \text{val(SDP-CF}^k)$. Moreover $(\text{val(SDP-CF}^k))_k$ is a monotone nonincreasing sequence because more constraints are added as k increases (so that the relaxations are tighter and tighter).

Restriction of the dual program

The program (D-CF) can be restricted so that we can derive a converging hierarchy of SDPs. It is essentially the minimisation of continuous functions for which we require additional constraints such as the fact that they are nonnegative. We will exploit the link between positive polynomials and sum-of-squares representation that is presented

in Appendix C.1. We recall the expression of the dual program (D-CF):

$$(\text{D-CF}) \begin{cases} \text{Find } (f_C)_{C \in \mathcal{M}} \in \prod_{C \in \mathcal{M}} C(O_C) \\[2ex] \text{minimising } \int_{O_C} f_C \; \mathrm{d} \, e_C \\[2ex] \text{subject to:} \\[1ex] \qquad \sum_{C \in \mathcal{M}} f_C \circ \rho_C^X \; \geq \; \mathbf{1}_{O_X} \\[1ex] \qquad \forall C \in \mathcal{M}, \; f_C \; \geq \; \mathbf{0}_{O_C} \, . \end{cases}$$

As this point we could derive the dual of program (SDP-CF^k) and show that this is indeed a restriction of the above program. For a more symmetric treatment, we restrict the dual program building on Appendix C.1 and Theorem 6. Instead of optimising over positive continuous functions, we restrict them to belong to the quadratic module $Q(g)$ and then further to $Q_k(g)$ for some $k \in \mathbb{N}_{>0}$. This requires that the degrees of SOS polynomials are fixed. For $k \in \mathbb{N}_{>0}$, we have

$$(\text{DSDP-CF}^k) \begin{cases} \text{Find } (p_C)_{C \in \mathcal{M}} \subset \Sigma^2 \mathbb{R}[\boldsymbol{x}]_k \text{ and } (\sigma_j)_{j=1,\dots,m} \subset \Sigma^2 \mathbb{R}[\boldsymbol{x}]_{k-1} \\[2ex] \text{maximising } \sum_{C \in \mathcal{M}} \int_{O_C} p_C \; \mathrm{d} \, e_C \\[2ex] \text{subject to:} \\[2ex] \qquad \sum_{C \in \mathcal{M}} p_C \circ \rho_C^X - \mathbf{1}_{O_X} = \sum_{j=0}^{m} \sigma_j g_j \, . \end{cases}$$

(DSDP-CF^k) is a restriction of (D-CF) so that for all $k \in \mathbb{N}_{>0}$, we have that $\mathsf{NCF}(e) = \mathrm{val}(\text{D-CF}) \leq \mathrm{val}(\text{SDP-CF}^k)$. Furthermore, $(\mathrm{val}(\text{SDP-CF}^k))_k$ is a monotone nonincreasing sequence.

Problems (SDP-CF^k) and (DSDP-CF^k) are indeed dual programs (see Proposition 5 in Appendix D).

8.2 Convergence of the hierarchy of SDPs

Finally, we prove that the constructed hierarchy provides a sequence of objective values that converges to the noncontextual fraction $\mathsf{NCF}(e)$.

Theorem 5. *The optimal values of the hierarchy of semidefinite programs (SDP-CF^k) (resp. (DSDP-CF^k)) provide monotonically decreasing upper bounds converging to the noncontextual fraction $\mathsf{NCF}(e)$ which is the value of (P-CF). That is*

$$\mathrm{val}(\text{SDP-CF}^k) \downarrow \mathrm{val}(\text{P-CF}) = \mathsf{NCF}(e) \quad \text{as } k \to \infty, \tag{32}$$

$$\mathrm{val}(\text{DSDP-CF}^k) \downarrow \mathrm{val}(\text{D-CF}) = \mathsf{NCF}(e) \quad \text{as } k \to \infty. \tag{33}$$

Proof. Because of the strong duality between the original infinite-dimensional linear programs we have

$$\mathrm{val}(\text{P-CF}) = \mathrm{val}(\text{D-CF}) = \mathsf{NCF}(e). \tag{34}$$

Moreover, for all $k \geq 1$ (SDP-CF^k) is a relaxation of (P-CF):

$$\text{val(SDP-CF}^k) \geq \text{val(P-CF)}. \tag{35}$$

And for all $k \geq 1$, (DSDP-CF^k) is a restriction of (D-CF):

$$\text{val(DSDP-CF}^k) \geq \text{val(D-CF)}. \tag{36}$$

Also for all $k \geq 1$, we have weak duality between (SDP-CF^k) and (DSDP-CF^k) (see Proposition 5):

$$\text{val(DSDP-CF}^k) \geq \text{val(SDP-CF}^k). \tag{37}$$

Thus for all $k \geq 1$:

$$\text{val(DSDP-CF}^k) \geq \text{val(SDP-CF}^k) \geq \mathsf{NCF}(e). \tag{38}$$

We already saw that $(\text{val(SDP-CF}^k))_k$ and $(\text{val(DSDP-CF}^k))_k$ form monotone nonincreasing sequences. We now show that $(\text{val(DSDP-CF}^k))_K$ converges to $\mathsf{NCF}(e)$. This is equivalent to showing that we can approximate any feasible solution[17] of program (D-CF) with a solution of (DSDP-CF^k) for a high enough rank k.

Fix $\varepsilon > 0$ and a feasible solution $(f_C)_{C \in \mathcal{M}} \in \prod_{C \in \mathcal{M}} C(O)$ of (D-CF). Then for all $C \in \mathcal{M}$, $f_C + \frac{\varepsilon}{|\mathcal{M}|}$ is a positive continuous function on O_C. Because O_C is compact (see Sect. 5.3) by the Stone–Weierstrass theorem, $f_C + \frac{\varepsilon}{|\mathcal{M}|}$ can be approximated by a positive polynomial. Thus there exist positive polynomials $p_C^\varepsilon \in \mathbb{R}[\boldsymbol{x}]$ such that for all contexts $C \in \mathcal{M}$ we have (in sup norm) that

$$\left\| f_C + \frac{\varepsilon}{|\mathcal{M}|} - p_C^\varepsilon \right\| \leq \frac{\varepsilon}{|\mathcal{M}|} \tag{39}$$

and also

$$\left\| \left(f_C + \frac{\varepsilon}{|\mathcal{M}|} - p_C^\varepsilon \right) \circ \rho_C^X \right\| < \frac{1}{|\mathcal{M}|} \min_{\boldsymbol{x} \in O_X} \left(\sum_{C \in \mathcal{M}} \left(f_C + \frac{\varepsilon}{|\mathcal{M}|} \right) \circ \rho_C^X(\boldsymbol{x}) - \boldsymbol{x} \right), \tag{40}$$

where the minimum is strictly positive as $\sum_C (f_C + \frac{\varepsilon}{|\mathcal{M}|}) \circ \rho_C^X > \mathbf{1}_{O_X}$.

From Eq. (39), the objective derived with $(p_C^\varepsilon)_C$ is ε-close to the original objective:

$$\left| \sum_{C \in \mathcal{M}} \int_{O_C} f_C \; \mathrm{d}e_C - \sum_{C \in \mathcal{M}} \int_{O_C} p_C^\varepsilon \; \mathrm{d}e_C \right| \leq \sum_{C \in \mathcal{M}} \int_{O_C} \left| f_C + \frac{\varepsilon}{|\mathcal{M}|} - p_C^\varepsilon \right| \; \mathrm{d}e_C \tag{41}$$

$$\leq \varepsilon . \tag{42}$$

Also from Eq. (40):

$$\sum_{C \in \mathcal{M}} p_C^\varepsilon \circ \rho_C^X - \mathbf{1} \tag{43}$$

---

[17] Note that program (D-CF) might not have an optimal solution in which case it only has an optimal solution in the closure of the feasible set. In that case, we can always find a sequence of feasible solutions converging to an optimal solution.

$$> \sum_{C \in \mathcal{M}} \left(f_C + \frac{\varepsilon}{|\mathcal{M}|} \right) \circ \rho_C^X - \min_{\boldsymbol{x} \in O_X} \left(\sum_{C \in \mathcal{M}} \left(f_C + \frac{\varepsilon}{|\mathcal{M}|} \right) \circ \rho_C^X(\boldsymbol{x}) - \boldsymbol{x} \right) - \mathbf{1}$$

(44)

$$\geq 0,$$

(45)

so that $\sum_{C \in \mathcal{M}} p_C^\varepsilon \circ \rho_C^X - \mathbf{1}_{O_X}$ is a positive polynomial on O_X. Next, because O_X is of the form required in Assumption 1, by Putinar's Positivellensatz (see Theorem 6), $\sum_{C \in \mathcal{M}} p_C^\varepsilon \circ \rho_C^X - \mathbf{1}_{O_X}$ belongs to the quadratic module $Q(g)$. Therefore, for a high enough rank $k \in \mathbb{N}$, it is a feasible solution of (DSDP-CF^k) and thus:

$$|\mathsf{NCF}(e) - \mathrm{val}(\text{DSDP-CF}^k)| \leq \varepsilon .$$

(46)

□

Outlook

Logical forms of contextuality, which are present at the level of the *possibilistic* rather than *probabilistic* information contained in an empirical model, remain to be considered (e.g. [2,8,38,62]). In the discrete setting, these can be treated by analysing 'possibilistic' empirical models obtained by considering the supports of the discrete-variable probability distributions [8], which indicate the elements of an outcome space that occur with non-zero probability. In general, the notion of support of a measure is not as straightforward, and the naïve approach is not viable since typically all singletons have measure 0. Nevertheless, supports can be defined in the setting of Borel measurable spaces, for instance, which in any case are the kind of spaces in which we are practically interested, in Sects. 5 and 8.

Approaches to contextuality that characterise obstructions to global sections using cohomology have had some success [5,11,27–29,65,66,75,77] and typically apply to logical forms of contextuality. An interesting prospect is to explore how the present framework may be employed to these ends, and to see whether the continuous-variable setting can open the door to new techniques that can be applied, or whether qualitatively new forms of contextual behaviour may be uncovered. A related direction to be developed is to understand how our treatment of contextuality can be further extended to continuous measurement spaces as proposed in [30].

Another direction to be explored is how our continuous-variable framework for contextuality can be extended to apply to more general notions of contextuality that relate not only to measurement contexts but also more broadly to contexts of preparations and transformations as well [63,81], noting that these also admit quantifiable relationships to quantum advantage [43,63].

Indeed, a major motivation to study contextuality is for its connections to quantum-over-classical advantages in informatic tasks. An important line of questioning is to ask what further connections can be found in the continuous-variable setting, and whether continuous-variable contextuality might offer advantages that outstrip those achievable with discrete-variable contextual resources. Note that it is known that infinite-dimensional quantum systems can offer certain additional advantages beyond finite-dimensional ones [80], though the empirical model that arises in that example is still a discrete-variable one in our sense.

The present work sets the theoretical basis for computational exploration of continuous-variable contextuality in quantum-mechanical empirical models. This, we hope, can provide new insights and inform other avenues to be developed in future work. It can also be useful in verifying the non-classicality of empirical models. Numerical implementation of the programs of Sect. 8 is of particular interest. The hierarchy of semi-definite programs can be used numerically to witness contextuality in continuous-variable experiments. Even if the time-complexity of the semi-definite program may increase drastically with its degree, a low-degree program can already provide a first witness of contextual behaviour.

Since our framework for continuous-variable contextuality is independent of quantum theory itself, it can equally be applied to 'empirical models' that arise in other, non-physical settings. The discrete-variable framework of [8] has led to a number of surprising connections and cross-fertilisations with other fields [3], including natural language [12], relational databases [1,16], logic [5,10,51], constraint satisfaction [7,9] and social systems [34]. It may be hoped that similar connections and applications can be found for the present framework to fields in which continuous-variable data is of central importance. For instance probability kernels of the kind we have used are also widely employed in machine learning (e.g. [45]), inviting intriguing questions about how our framework might be used or what advantages contextuality may confer in that setting.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Appendices

A. Linear Programs in Standard Form

This appendix may be of particular interest to readers familiar with global optimisation. We express the problems (P-CF) and (D-CF) in the standard form of infinite linear programming [17, IV–(7.1)]. We recall them below:

$$
\text{(P-CF)} \left\{
\begin{array}{l}
\text{Find } \mu \in \mathbb{M}_+(\boldsymbol{O}_X) \\
\text{maximising } \mu(O_X) \\
\text{subject to:} \\
\quad \forall C \in \mathcal{M}, \ \mu|_C \ \leq \ e_C \\
\quad \mu \ \geq \ 0 \,.
\end{array}
\right.
$$

$$
\text{(D-CF)} \begin{cases}
\text{Find } (f_C)_{C \in \mathcal{M}} \in \prod_{C \in \mathcal{M}} C(O_C) \\[2ex]
\text{minimising } \displaystyle\int_{O_C} f_C \ d\, e_C \\[2ex]
\text{subject to:} \\[1ex]
\qquad \displaystyle\sum_{C \in \mathcal{M}} f_C \circ \rho_C^X \geq \mathbf{1}_{O_X} \\[2ex]
\qquad \forall C \in \mathcal{M}, \ f_C \geq \mathbf{0}_{O_C} .
\end{cases}
$$

Problems (P-CF) and (D-CF) are indeed infinite linear programs as both the objective and the constraints are linear with respect to the unknown measure $\mu \in \mathbb{M}_{\pm}(\boldsymbol{O}_X)$. To write (P-CF) in the standard form [17], we introduce the following spaces:

- $E_1 := \mathbb{M}_{\pm}(\boldsymbol{O}_X)$.
- $F_1 := C(O_X)$, the dual space of E_1.
- $E_2 := \displaystyle\prod_{C \in \mathcal{M}} \mathbb{M}_{\pm}(\boldsymbol{O}_C)$.
- $F_2 := \displaystyle\prod_{C \in \mathcal{M}} C(O_C)$, the dual space of E_2.

The dualities $\langle -, - \rangle_1 : E_1 \times F_1 \longrightarrow \mathbb{R}$ and $\langle -, - \rangle_2 : E_2 \times F_2 \longrightarrow \mathbb{R}$ are defined as follows:

$$
\forall \mu \in E_1, \ f \in F_1, \qquad \langle \mu, f \rangle_1 := \int_{O_X} f \ d\mu
$$

$$
\forall (\nu_C) \in E_2, \ (f_C) \in F_2, \qquad \langle (\nu_C), (f_C) \rangle_2 := \sum_{C \in \mathcal{M}} \int_{O_C} f_C \ d\nu_C ,
$$

where, for simplicity, we have omitted $C \in \mathcal{M}$ as a subscript for the families of functions. We fix K_1 to be the convex cone of positive measures in $E_1 = \mathbb{M}_{\pm}(\boldsymbol{O}_C)$ and K_2 to be the convex cone of families of positive measures in $E_2 = \prod_{C \in \mathcal{M}} \mathbb{M}_{\pm}(\boldsymbol{O}_C)$. Then K_1^* is the convex cone of positive function in $F_1 = C(O_X)$ and K_2^* is the convex cone of families of positive functions in $F_2 = \prod_{C \in \mathcal{M}} C(O_C)$.

Let $A : E_1 \longrightarrow E_2$ be the following linear transformation: for $\mu \in E_1$

$$
A(\mu) := (\mu|_C)_{C \in \mathcal{M}} \in E_2 .
$$

We also define the linear transformation $A^* : F_2 \longrightarrow F_1$ as follows: for $(f_C) \in F_2$,

$$
A^*((f_C)) := \sum_{C \in \mathcal{M}} f_C \circ \rho_C^X \in F_1 .
$$

We can verify that A^* is the dual transformation of A: given $\mu \in E_1$ and $(f_C) \in F_2$, we have

$$
\langle A(\mu), (f_C) \rangle_2 = \langle (\mu|_C), (f_C) \rangle_2 \tag{47}
$$

$$
= \sum_{C \in \mathcal{M}} \int_{O_C} f_C \ d\mu|_C \tag{48}
$$

$$= \int_{O_X} \sum_{C \in \mathcal{M}} f_C \circ \rho_C^X \, d\mu \tag{49}$$

$$= \langle \mu, \sum_{C \in \mathcal{M}} f_C \circ \rho_C^X \rangle_1 \tag{50}$$

$$= \langle \mu, A^*((f_C)) \rangle_1 . \tag{51}$$

Now fixing the vector function in the objective to be $c := -\mathbf{1}_{O_X} \in F_1$ and the vector in the constraints to be $b := (-e_C)_{C \in \mathcal{M}} \in E_2$, the program (P-CF) (resp. (D-CF)) can be expressed as in the standard form given in [17]:

$$\text{LP} \begin{cases} \text{Find } e_1 \in E_1 \\ \text{minimising } \langle e_1, c \rangle_1 \\ \text{subject to:} \\ A(e_1) \geq_{K_2} b \\ e_1 \geq_{K_1} 0 . \end{cases}$$

$$(\text{D-LP}) \begin{cases} \text{Find } f_2 \in F_2 \\ \text{maximising } \langle b, f_2 \rangle_2 \\ \text{subject to:} \\ A^*(f_2) \leq_{K_1^*} c \\ f_2 \geq_{K_2} 0 . \end{cases}$$

Note that the minus sign in the vectors c and b was added because we chose the primal program in the standard form to be a minimisation problem while the primal program (P-CF) at hand is a maximisation problem.

B. Proof of Proposition 2: Zero Duality Gap

In this appendix we give a full proof of Proposition 2; i.e. that strong duality holds between problems (P-CF) and (D-CF).

Proof. To show strong duality, we rely on [17, Theorem 7.2]. Because $\mu_0 = \mathbf{0}_{O_X}$—the measure that assigns 0 to every measurable set of O_X—is a feasible solution for (P-CF) and the noncontextual fraction lies between 0 and 1, (P-CF) is consistent with finite value. Thus it suffices to show that the following cone

$$\mathcal{K} = \{(A(\mu), \langle \mu, c \rangle_1) \mid \mu \in K_1\} = \{((\mu|_C)_C, \mu(O_X)) \mid \mu \in K_1\} \tag{52}$$

is weakly closed in $E_2 \oplus \mathbb{R}$ (i.e. closed in the weak topology of K_1) where we recall that K_1 is the convex cone of positive measures in $E_1 = \mathbb{M}_\pm(O_X)$.

We first notice that the linear transformation A is a bounded linear operator and thus continuous. Boundedness comes from the fact that for all $\mu \in K_1$,

$$\|A(\mu)\|_{E_2} = \|(\mu|_C)_C\|_{E_2} \tag{53}$$

$$= \sum_{C \in \mathcal{M}} \|\mu|_C\|_{\mathbb{M}_\pm(O_C)} \tag{54}$$

$$\leq \sum_{C \in \mathcal{M}} \|\mu\|_{E_1} \tag{55}$$

$$= |\mathcal{M}| \, \|\mu\|_{E_1} \, , \tag{56}$$

where we take the strong topology—i.e. the norm induced by the total variation distance—on finite-signed measure spaces. It is defined as:

$$\|\mu\|_{\mathbb{M}_{\pm}(U)} = |\mu|(U) \, .$$

We also equip the finite product space $E_2 = \prod_{C \in \mathcal{M}} \mathbb{M}_{\pm}(O_C)$ with the norm obtained by summing[18] the individual total variation norms. The inequality in Eq. (55) is due to the fact that $\mu \in K_1$ so this is a positive measure and thus $\|\mu|_C\|_{\mathbb{M}_{\pm}(O_C)} \leq \|\mu\|_{E_1}$. This, of course, extends to the weak topology.

Secondly, we consider a sequence $(\mu^k)_{k \in \mathbb{N}}$ in K_1 and we want to show that the accumulation point $((\Theta_C)_C, \lambda) = \lim_{k \to \infty} \left(A(\mu^k), \langle \mu^k, c \rangle_1 \right)$ belongs to $\mathcal{K}$, where $\Theta = (\Theta_C)_C \in E_2$ and $\lambda \in \mathbb{R}$. If we consider the product of indicator functions $(\mathbf{1}_{O_C})_C \in F_2$ then $\langle A(\mu^k), (\mathbf{1}_{O_C}) \rangle_2 = \sum_{C \in \mathcal{M}} \mu|_C^k(O_C) \longrightarrow_k \sum_{C \in \mathcal{M}} \Theta_C(O_C) < \infty$ as Θ_C is a finite measure for all maximal contexts $C \in \mathcal{M}$. Then, because $\mathcal{M}$ is a covering family of X, we have that, for all $k \in \mathbb{N}$ $\mu^k(O_X) \leq \sum_{C \in \mathcal{M}} \mu^k|_C(O_C) < \infty$. Since $(\mu^k) \in K_1^{\mathbb{N}}$ is a sequence of positive measures, this implies that (μ^k) is bounded. Next, by weak-$*$ compactness of the unit ball (Alaoglu's theorem [58]), there exists a subsequence $(\mu^{k_i})_i$ that converges weakly to an element $\omega \in K_1$. By continuity of A, we obtain that the accumulation point is such that $((\Theta_C)_C, \lambda) = (A(\omega), \langle \omega, c \rangle_1) \in \mathcal{K}$. $\qquad \square$

C. The Lasserre–Parrilo Hierarchy

Below we introduce the Lasserre–Parrilo hierarchy for relaxing infinite-dimensional linear programs known as Generalised Moment Problems [53,54,69]. We start by giving insightful results: Sect. C.1 provides results concerning the representation of positive polynomials while Sect. C.2 provides results to understand when a sequence can be represented by a Borel measure.

Notation, terminology
We work in $\mathbb{R}^d$ for $d \in \mathbb{N}^*$. We fix K to be a generic Borel measurable subspace of $\mathbb{R}^d$. Below we fix some multi-index notations. Let $\mathbb{R}[\boldsymbol{x}]$ denote the ring of real polynomials in the variables $\boldsymbol{x} \in \mathbb{R}^d$, and let $\mathbb{R}[\boldsymbol{x}]_k \subset \mathbb{R}[\boldsymbol{x}]$ contain those polynomials of total degree at most k. The latter forms a vector space of dimension $s(k) := \binom{d+k}{k}$, with a canonical basis consisting of monomials $\boldsymbol{x}^{\boldsymbol{\alpha}} := x_1^{\alpha_1} \cdots x_d^{\alpha_d}$ indexed by the set $\mathbb{N}_k^d := \{ \boldsymbol{\alpha} \in \mathbb{N}^d \mid |\boldsymbol{\alpha}| \leq k \}$ where $|\boldsymbol{\alpha}| := \sum_{i=1}^d \alpha_i$. For $k \in \mathbb{N}$, $\boldsymbol{x} \in \mathbb{R}^d$, we define $\mathbf{v}_k(\boldsymbol{x}) := (\boldsymbol{x}^{\boldsymbol{\alpha}})_{|\boldsymbol{\alpha}| \leq d} = (1, x_1, \ldots, x_n, x_1^2, x_1 x_2, \ldots, x_n^k)^T$ the vector of monomials of total degree less or equal than k.

Any $p \in \mathbb{R}[\boldsymbol{x}]_k$ is associated with a vector of coefficients $\boldsymbol{p} := (p_{\boldsymbol{\alpha}}) \in \mathbb{R}^{s(k)}$ via expansion in the canonical basis as $p(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha} \in \mathbb{N}_k^d} p_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}}$.

---

[18] Categorically, this is a coproduct.

Moment problem in probability

Given a finite set of indices Γ, a set of reals $\{\gamma_j : j \in \Gamma\}$ and functions $h_j \colon K \longrightarrow \mathbb{R}$, $j \in \Gamma$, that are integrable with respect to every measure $\mu \in \mathbb{M}_\pm(K)$, the corresponding *Global Moment Problem* (GMP) can be expressed as:

$$
(\text{GMP}) \begin{cases} \text{Find } \mu \in \mathbb{M}_\pm(K) \\ \text{maximising } \mu(K) \\ \text{subject to:} \\ \qquad \forall j \in \Gamma, \quad \int_K h_j \, \mathrm{d}\mu \le \gamma_j \, . \end{cases}
$$

It dual program can be expressed as:

$$
(\text{D-GMP}) \begin{cases} \text{Find } \boldsymbol{\lambda} \in \mathbb{R}^\Gamma \\ \text{minimising } \displaystyle\sum_{j\in\Gamma} \gamma_j \lambda_j \\ \text{subject to:} \\ \qquad \forall \boldsymbol{x} \in K, \quad \displaystyle\sum_{j\in\Gamma} \lambda_j h_j(\boldsymbol{x}) - \boldsymbol{x} \ge 0 \\ \qquad \forall j \in \Gamma, \quad \lambda_j \ge 0 \, . \end{cases}
$$

C.1. Positive polynomials and sum-of-squares

Here we present the link between positive polynomials and sum-of-squares representation so that we can derive a converging hierarchy of restriction problems for program (D-GMP).

Definition 14. A polynomial $p \in \mathbb{R}[\boldsymbol{x}]$ is a *sum-of-squares (SOS)* polynomial if there exists a finite family of polynomials $(q_i)_{i\in I}$ such that $p = \sum_{i\in I} q_i^2$.

SOS polynomials are widely used in convex optimisation. We will denote by $\Sigma^2\mathbb{R}[\boldsymbol{x}] \subset \mathbb{R}[\boldsymbol{x}]$ the set of (multivariate) SOS polynomials, and $\Sigma^2\mathbb{R}[\boldsymbol{x}]_k \subset \Sigma^2\mathbb{R}[\boldsymbol{x}]$ the set of SOS polynomials of degree at most $2k$. The following proposition hints towards the reason why it is desirable to be able to look for a sum-of-squares decomposition: it can be cast as a semidefinite optimisation problem.

Proposition 3 (Prop. 2.1, [54]). *A polynomial $p \in \mathbb{R}[\boldsymbol{x}]_{2k}$ has a sum-of-squares decomposition if and only if there exists a real symmetric positive semidefinite matrix $Q \in \mathsf{Sym}_{s(k)}$ such that $\forall \boldsymbol{x} \in \mathbb{R}^d, \ p(\boldsymbol{x}) = \mathbf{v}_k(\boldsymbol{x})^T Q \mathbf{v}_k(\boldsymbol{x})$.*

Then we will be looking at conditions under which a nonnegative polynomial can be expressed as a sum-of-squares polynomial. This is in essence the question raised by Hilbert in his 17th conjecture [44].

Definition 15. For a family $q = (q_j)_{j\in\{1,...,m\}}$ of polynomials, the set:

$$
Q(q) := \left\{ \sum_{j=0}^m \sigma_j q_j \mid (\sigma_j)_{j\in\{0,...,m\}} \subset \Sigma^2\mathbb{R}[\boldsymbol{x}] \right\} \tag{57}
$$

is a convex cone in $\mathbb{R}[x]$ called the *quadratic module* generated by the family q with, for convenience, $q_0 = 1$ added. For $k \in \mathbb{N}$, we define $Q_k(q)$ to be the quadratic module $Q(q)$ where we further impose that $(\sigma_j)_{j \in \{0,\dots,m\}} \subset \Sigma^2\mathbb{R}[x]_k$ i.e. we limit the degree of SOS polynomials.

In the family of polynomials $(g_j)_{j \in \{1,\dots,m\}}$ we add $g_0 = 1$ for convenience.

Assumptions 1. Let $K \subset \mathbb{R}^d$. We make the following three assumptions on K.

  (i) Suppose K is a basic semi-algebraic set i.e. there exists a family of polynomials
  $g = (g_j)_{j \in \{1,\dots,m\}} \in \mathbb{R}[x]^m$ of degrees $\deg(g_j)$ respectively such that:

$$K := \left\{ x \in \mathbb{R}^d \mid \forall j = 1, \dots, m, \ g_j(x) \geq 0 \right\}. \tag{58}$$

 (ii) Further suppose that K is compact.
(iii) Finally suppose that there exists $u \in Q(q)$ such that the level set $\left\{ x \in \mathbb{R}^d \mid u(x) \geq 0 \right\}$
     is compact.

The following theorem is the key result that we will exploit for deriving the hierarchy of SDP restrictions for the dual program (GMP).

Theorem 6 (Putinar's Positivellensatz [73]). *Let $K \subset \mathbb{R}^d$ satisfy Assumptions 1. If $p \in \mathbb{R}[x]$ is strictly positive on K then $p \in Q(g)$, that is*

$$p = \sum_{j=0}^{m} \sigma_j g_j \tag{59}$$

for some sum-of-squares polynomials $\sigma_j \in \Sigma^2\mathbb{R}[x]$ for $j = 0, 1, \dots, m$.

A proof can also be found in [56].

Using the results above and Assumption 1, one can derive a hierarchy of SDPs [54] which provide a converging sequence of optimal values towards the value of program (D-GMP):

D-(GMP^k)
$$\begin{cases} \text{Find } \lambda = (\lambda_j)_{j \in \Gamma} \in \mathbb{R}^\Gamma \text{ and } \forall j = 0, \dots, m, \ f_j \in \Sigma^2\mathbb{R}[x]_{k - \lceil \frac{\deg(g_j)}{2} \rceil} \\[2mm] \text{minimising } y_0 \\[1mm] \text{subject to:} \\[2mm] \sum_{j \in \Gamma} \lambda_j h_j - \mathbf{1}_K = \sum_{j=0}^{m} f_j g_j \\[2mm] \forall j \in \Gamma, \quad \lambda_j \geq 0 \,. \end{cases}$$

C.2. Moment sequences and moment matrices

In this subsection, we want to understand why the program (P-CF) can be relaxed so that a converging hierarchy of SDPs can be derived. The program (P-CF) is essentially a maximisation problem on finite-signed Borel measures with additional constraints such as the fact that these are proper measures (i.e. they are nonnegative). We will represent a measure by its moment sequence and find conditions for which this moment sequence has a (unique) representing Borel measure.

Definition 16. Given a sequence $\boldsymbol{y} = (y_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha} \in \mathbb{N}^d} \in \mathbb{R}^{\mathbb{N}^d}$, we define the linear functional $L_{\boldsymbol{y}} \colon \mathbb{R}[\boldsymbol{x}] \longrightarrow \mathbb{R}$ by

$$L_{\boldsymbol{y}}(p) := \sum_{\boldsymbol{\alpha} \in \mathbb{N}^d} p_{\boldsymbol{\alpha}} y_{\boldsymbol{\alpha}}. \tag{60}$$

Definition 17. Given a measure $\mu \in \mathbb{M}(K)$, its *moment sequence* $\boldsymbol{y} = (y_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha} \in \mathbb{N}^d} \in \mathbb{R}^{\mathbb{N}^d}$ is given by

$$y_{\boldsymbol{\alpha}} := \int_K \boldsymbol{x}^{\boldsymbol{\alpha}} \, \mathrm{d}\mu(\boldsymbol{x}) . \tag{61}$$

We say that $\boldsymbol{y}$ has a unique representing measure μ when there exists a unique μ such that Eq. (61) holds. If μ is unique then we say it is determinate (i.e. determined by its moments).

The linear functional $L_{\boldsymbol{y}}$ then gives integration of polynomials with respect to μ i.e. for any $p \in \mathbb{R}[\boldsymbol{x}]$:

$$\begin{aligned}
L_{\boldsymbol{y}}(p) = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^d} p_{\boldsymbol{\alpha}} y_{\boldsymbol{\alpha}} &= \sum_{\boldsymbol{\alpha} \in \mathbb{N}^d} p_{\boldsymbol{\alpha}} \int_K \boldsymbol{x}^{\boldsymbol{\alpha}} \, \mathrm{d}\mu(\boldsymbol{x}) = \int_K \sum_{\boldsymbol{\alpha} \in \mathbb{N}^d} p_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}} \, \mathrm{d}\mu(\boldsymbol{x}) \\
&= \int_K p(\boldsymbol{x}) \, \mathrm{d}\mu(\boldsymbol{x}) \\
&= \int_K p \, \mathrm{d}\mu,
\end{aligned} \tag{62}$$

where we reversed summation and integration because the sum is finite since p is a polynomial.

The following theorem is often used in optimisation theory over measures as it provides a necessary and sufficient condition for a sequence to have a representing measure.

Theorem 7 (Riesz-Haviland [41]). *Let* $\boldsymbol{y} = (y_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha} \in \mathbb{N}^d} \in \mathbb{R}^{\mathbb{N}^d}$ *and suppose that* $K \subseteq \mathbb{R}^d$ *is closed. Then* $\boldsymbol{y}$ *has a representation (nonnegative) measure i.e. there exists* μ *a measure on* K *such that:*

$$\forall \boldsymbol{\alpha} \in \mathbb{N}^d, \ \int_K \boldsymbol{x}^{\boldsymbol{\alpha}} \, \mathrm{d}\mu = y_{\boldsymbol{\alpha}}$$

if and only if $L_{\boldsymbol{y}}(p) \geq 0$ *for all polynomials* $p \in \mathbb{R}[\boldsymbol{x}]$ *nonnegative on* K.

We recall that for $k \in \mathbb{N}$, $s(k) = \binom{d+k}{k}$.

Definition 18. For each $k \in \mathbb{N}$, the *moment matrix* of order k, $M_k(\boldsymbol{y}) \in \mathsf{Sym}_{s(k)}$, of a truncated sequence $(y_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha} \in \mathbb{N}^d_{2k}}$ is the $s(k) \times s(k)$ symmetric matrix with rows and columns indexed by $\mathbb{N}^d_k$ (i.e. by the canonical basis for $\mathbb{R}[\boldsymbol{x}]_k$) defined as follows: for any $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{N}^d_k$,

$$(M_k(\boldsymbol{y}))_{\boldsymbol{\alpha}, \boldsymbol{\beta}} := L_{\boldsymbol{y}}(\boldsymbol{x}^{\boldsymbol{\alpha}+\boldsymbol{\beta}}) = y_{\boldsymbol{\alpha}+\boldsymbol{\beta}} . \tag{63}$$

Definition 19. Given a polynomial $p \in \mathbb{R}[\boldsymbol{x}]$, the *localising matrix* $M_k(p\boldsymbol{y}) \in \mathsf{Mat}_{s(k)}(\mathbb{R})$ of a moment sequence $(y_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha} \in \mathbb{N}^d} \in \mathbb{R}^{\mathbb{N}^d}$ is defined by: for all $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{N}_k^d$,

$$(M_k(p\boldsymbol{y}))_{\boldsymbol{\alpha},\boldsymbol{\beta}} := L_{\boldsymbol{y}}(p(x)x^{\boldsymbol{\alpha}+\boldsymbol{\beta}}) = \sum_{\boldsymbol{\gamma} \in \mathbb{N}^d} p_{\boldsymbol{\gamma}} \, y_{\boldsymbol{\alpha}+\boldsymbol{\beta}+\boldsymbol{\gamma}} \, . \tag{64}$$

The localising matrix reduces to the moment matrix for $p = 1$. For well-defined moment sequences, i.e. sequences that have a representing finite Borel measure, moment matrices and localising matrices are positive semidefinite, which provides insight on the reason why problem (P-CF) can be relaxed to a problem with positive semidefiniteness constraints.

Proposition 4. *Let $\boldsymbol{y} = (y_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha} \in \mathbb{N}^d} \in \mathbb{R}^{\mathbb{N}^d}$ be a sequence of moments for some finite Borel measure μ on $\boldsymbol{K}$. Then for all $k \in \mathbb{N}$, $M_k(\boldsymbol{y}) \succeq 0$. If μ has support contained in the set $\{\boldsymbol{x} \in \boldsymbol{K} \mid g(\boldsymbol{x}) \geq 0\}$ for some polynomial $g \in \mathbb{R}[\boldsymbol{x}]$ then, for all $k \in \mathbb{N}$, $M_k(g\boldsymbol{y}) \succeq 0$.*

Proof. Let $\boldsymbol{y} = (y_{\boldsymbol{\alpha}})$ be the moment sequence of a given Borel measure μ on $\boldsymbol{K}$. Fix $k \in \mathbb{N}$. For any vector $\boldsymbol{v} \in \mathbb{R}^{s(k)}$ (noting that $\boldsymbol{v}$ is canonically associated with a polynomial $v \in \mathbb{R}[\boldsymbol{x}]_k$ in the basis $(\boldsymbol{x}^{\boldsymbol{\alpha}})$):

$$\boldsymbol{v}^T M_k(\boldsymbol{y})\boldsymbol{v} = \sum_{\boldsymbol{\alpha},\boldsymbol{\beta} \in \mathbb{N}_k^d} v_{\boldsymbol{\alpha}} \, y_{\boldsymbol{\alpha}+\boldsymbol{\beta}} \, v_{\boldsymbol{\beta}} \tag{65}$$

$$= \sum_{\boldsymbol{\alpha},\boldsymbol{\beta} \in \mathbb{N}_k^d} v_{\boldsymbol{\alpha}} v_{\boldsymbol{\beta}} \int_{\boldsymbol{K}} x^{\boldsymbol{\alpha}+\boldsymbol{\beta}} \, \mathrm{d}\mu \tag{66}$$

$$= \int_{\boldsymbol{K}} \left(\sum_{\boldsymbol{\alpha} \in \mathbb{N}_k^d} v_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}} \right)^2 \, \mathrm{d}\mu \tag{67}$$

$$= \int_{\boldsymbol{K}} v^2(\boldsymbol{x}) \, \mathrm{d}\mu \geq 0 \, . \tag{68}$$

Thus $M_k(\boldsymbol{y}) \succeq 0$.

Similarly we can prove that the localising matrix $M_k(g\boldsymbol{y})$ is positive semidefinite when g is a nonnegative polynomial on the support of μ. Indeed for all $\boldsymbol{v} \in \mathbb{R}^{s(k)}$:

$$\boldsymbol{v}^T M_k(g\boldsymbol{y})\boldsymbol{v} = \int_{\boldsymbol{K}} v^2(\boldsymbol{x})g(\boldsymbol{x}) \, \mathrm{d}\mu \geq 0 \, , \tag{69}$$

which concludes the proof.   □

The following theorem, which is the dual facet of Theorem 6, is the key result for deriving the hierarchy of SDP relaxations for the primal problem (P-CF). It provides a necessary and sufficient condition for a sequence to have a representing measure.

Theorem 8 (Theorem 3.8 [54]). *Let $\boldsymbol{y} = (y_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha} \in \mathbb{N}^d} \in \mathbb{R}^{\mathbb{N}^d}$ be a given infinite sequence in $\mathbb{R}$. Let $K \subset \mathbb{R}^d$ satisfy Assumptions 1. Then $\boldsymbol{y}$ has a finite Borel representing measure with support contained in K if and only if:*

$$M_k(\boldsymbol{y}) \succeq 0, \; \forall k \in \mathbb{N}, \tag{70}$$
$$M_k(g_j \boldsymbol{y}) \succeq 0, \; \forall j = 1, \ldots, m, \; \forall k \in \mathbb{N}. \tag{71}$$

Using the results above and Assumption 1, one can derive a hierarchy of SDPs [54] which provide a converging sequence of optimal values towards the value of program (GMP):

$$(\text{GMP}^k) \begin{cases} \text{Find } \boldsymbol{y} = (y_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha} \in \mathbb{N}_{2k}^d} \in \mathbb{R}^{s(2k)} \\[4pt] \text{maximising } y_0 \\[4pt] \text{subject to:} \\[4pt] \forall j \in \Gamma, \quad L_{\boldsymbol{y}}(h_j) \le \gamma_j \\[4pt] M_k(\boldsymbol{y}) \succeq 0 \\[4pt] \forall i \in 1, \dots, m, \quad M_{k - \lceil \frac{\deg(g_i)}{2} \rceil}(g_i \, \boldsymbol{y}) \succeq 0 . \end{cases}$$

We refer readers to [54] for the proof of convergence of the hierarchies given by programs (GMP^k) and (D-GMP^k).

D. Duality between programs (SDP-CF^k) and (DSDP-CF^k)

As mentioned above, we chose to derive programs (SDP-CF^k) and (DSDP-CF^k) using dual arguments. These programs should therefore be dual to one another, which will immediately provide weak duality. We prove this for completeness.

Proposition 5. *The program* (DSDP-CF^k) *is the dual formulation of the program* (SDP-CF^k).

Proof. We start by rewriting $M_k(\boldsymbol{y})$ as $\sum_{\boldsymbol{\alpha} \in \mathbb{N}_k^d} y_{\boldsymbol{\alpha}} A_{\boldsymbol{\alpha}}$ and $M_{k-1}(g_j \, \boldsymbol{y})$ as $\sum_{\boldsymbol{\alpha} \in \mathbb{N}_k^d} y_{\boldsymbol{\alpha}} B_{\boldsymbol{\alpha}}^j$ for $1 \le j \le m$ and for appropriate real symmetric matrices $A_{\boldsymbol{\alpha}}$ and $(B_{\boldsymbol{\alpha}}^j)_j$. For instance, in the basis $(\boldsymbol{x}^{\boldsymbol{\alpha}})$:

$$(A_{\boldsymbol{\alpha}})_{s,t} = \left(\begin{cases} 1 \text{ if } s + t = \boldsymbol{\alpha} \\ 0 \text{ otherwise} \end{cases} \right)_{s,t} .$$

From $A_{\boldsymbol{\alpha}}$, we also extract $A_{\boldsymbol{\alpha}}^C$ for $C \in \mathcal{M}$ in order to rewrite $M_k(\boldsymbol{y}|_C)$ as $\sum_{\boldsymbol{\alpha} \in \mathbb{N}_k^d} y_{\boldsymbol{\alpha}} A_{\boldsymbol{\alpha}}^C$. This amounts to identifying which matrices $(A_{\boldsymbol{\alpha}})$ contribute to a given context $C \in \mathcal{M}$. Then (SDP-CF^k) can be rewritten as:

$$(\text{SDP-CF}^k) \begin{cases} \text{Find } \boldsymbol{y} \in \mathbb{R}^{s(2k)} \\[4pt] \text{maximising } y_0 \\[4pt] \text{subject to:} \\[4pt] \forall C \in \mathcal{M}, \; M_k(\boldsymbol{y}^{e,C}) - \sum_{\boldsymbol{\alpha} \in \mathbb{N}_k^d} y_{\boldsymbol{\alpha}} A_{\boldsymbol{\alpha}}^C \succeq 0, \\[10pt] \forall j = 1, \dots, m, \; \sum_{\boldsymbol{\alpha} \in \mathbb{N}_k^d} y_{\boldsymbol{\alpha}} B_{\boldsymbol{\alpha}}^j \succeq 0, \\[10pt] \sum_{\boldsymbol{\alpha} \in \mathbb{N}_k^d} y_{\boldsymbol{\alpha}} A_{\boldsymbol{\alpha}} \succeq 0 . \end{cases}$$

Via the Lagrangian method, this is equivalent to:

$$\sup_{\substack{\mathbf{y} \in \mathbb{R}^{s(k)}}} \inf_{\substack{(X^C),(Y_j),Z \\ \text{SDP matrices}}} \mathcal{L}(\mathbf{y}, (X_C), Y, (Z_j)), \tag{72}$$

with

$$\begin{aligned} \mathcal{L}(\mathbf{y}, (X^C), (Y_j), Z) = {} & y_{\mathbf{0}} \\ & + \sum_{C \in \mathcal{M}} \text{Tr}(M_k(\mathbf{y}^{e,C})X^C) - \sum_{C \in \mathcal{M}} \sum_{\boldsymbol{\alpha} \in \mathbb{N}_k^d} y_{\boldsymbol{\alpha}} \text{Tr}(A_{\boldsymbol{\alpha}}^C X^C) \\ & + \sum_{j=1}^{m} \sum_{\boldsymbol{\alpha} \in \mathbb{N}_k^d} y_{\boldsymbol{\alpha}} \text{Tr}(B_{\boldsymbol{\alpha}}^j Y_j) \\ & + \sum_{\boldsymbol{\alpha} \in \mathbb{N}_k^d} y_{\boldsymbol{\alpha}} \text{Tr}(A_{\boldsymbol{\alpha}} Z) \;. \end{aligned} \tag{73}$$

The dual program corresponds to

$$\inf_{\substack{(X^C),(Y_j),Z \\ \text{SDP matrices}}} \sup_{\mathbf{y} \in \mathbb{R}^{s(k)}} \mathcal{L}(\mathbf{y}, (X^C), (Y_j), Z) \;. \tag{74}$$

We rewrite the Lagrangian as:

$$\begin{aligned} \mathcal{L}(\mathbf{y}, (X_C), Y, (Z_j)) = {} & \sum_{C \in \mathcal{M}} \text{Tr}(M_k(\mathbf{y}^{e,C})X^C) \\ & + \sum_{\boldsymbol{\alpha} \in \mathbb{N}_k^d} y_{\boldsymbol{\alpha}} \left(\delta_{\boldsymbol{\alpha},\mathbf{0}} - \sum_{C \in \mathcal{M}} \text{Tr}(A_{\boldsymbol{\alpha}}^C X^C) + \sum_{j=1}^{m} \text{Tr}(B_{\boldsymbol{\alpha}}^j Y_j) + \text{Tr}(A_{\boldsymbol{\alpha}} Z) \right) \;. \end{aligned} \tag{75}$$

Then the dual program of (SDP-CF^k) reads:

$$\left\{ \begin{aligned} & \text{Find } (X^C)_{C \in \mathcal{M}}, (Y_j)_{j=1,\dots,m} \text{ and } Z \text{ SDP matrices} \\ & \text{maximising } \sum_{C \in \mathcal{M}} \text{Tr}(M_k(\mathbf{y}^{e,C})X^C) \\ & \text{subject to:} \\ & \forall \boldsymbol{\alpha} \in \mathbb{N}_k^d, \; \sum_{C \in \mathcal{M}} \text{Tr}(A_{\boldsymbol{\alpha}}^C X^C) - \sum_{j=1}^{m} \text{Tr}(B_{\boldsymbol{\alpha}}^j Y_j) - \text{Tr}(A_{\boldsymbol{\alpha}} Z) = \delta_{\boldsymbol{\alpha},\mathbf{0}} \;. \end{aligned} \right.$$

We finally show that the above program exactly corresponds to (DSDP-CF^k). We start with the objective. For all $C \in \mathcal{M}$, with X^C a positive semidefinite matrix:

$$\text{Tr}(M_k(\mathbf{y}^{e,C})X^C) = \sum_{\boldsymbol{\alpha}} \sum_{\boldsymbol{\beta}} (M_k(\mathbf{y}^{e,C}))_{\boldsymbol{\alpha}\boldsymbol{\beta}} (X^C)_{\boldsymbol{\beta}\boldsymbol{\alpha}} \tag{76}$$

$$= \sum_{\boldsymbol{\alpha}} \sum_{\boldsymbol{\beta}} y_{\boldsymbol{\alpha}+\boldsymbol{\beta}}^{e,C} (X^C)_{\boldsymbol{\alpha}\boldsymbol{\beta}} \tag{77}$$

$$= \sum_{\boldsymbol{\alpha}} \sum_{\boldsymbol{\beta}} \int_{\mathcal{O}_C} x^{\boldsymbol{\alpha}+\boldsymbol{\beta}} \; \mathrm{d}\, e_C (X^C)_{\boldsymbol{\alpha}\boldsymbol{\beta}} \tag{78}$$

$$= \int_{\mathcal{O}_C} \mathbf{v}_k(\boldsymbol{x})^T X^C \mathbf{v}_k(\boldsymbol{x}) \; \mathrm{d}\, e_C \tag{79}$$

$$= \int_{\mathcal{O}_C} p_C \; \mathrm{d}\, e_C, \tag{80}$$

with for all $C \in \mathcal{M}$, $p_C \in \Sigma^2 \mathbb{R}[\boldsymbol{x}]_k$ a sum-of-squares polynomial via Proposition 3 and where we used $\mathbf{v}_k(\boldsymbol{x})$ the vectors of monomials of maximal total degree k.
Now, to retrieve the constraint, we multiply each side by $\boldsymbol{x}^{\boldsymbol{\alpha}}$ and we sum for all $\boldsymbol{\alpha}$:

$$\sum_{C \in \mathcal{M}} \mathrm{Tr}(\sum_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}} A^C_{\boldsymbol{\alpha}} X^C) - \mathbf{1} = \sum_{j=1}^{m} \mathrm{Tr}(\sum_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}} B^j_{\boldsymbol{\alpha}} Y_j) + \mathrm{Tr}(\sum_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}} A_{\boldsymbol{\alpha}} Z) \tag{81}$$

Recalling the definition of moment and localising matrices:

$$\sum_{\boldsymbol{\alpha}} A^C_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}} = \mathbf{v}_k(\boldsymbol{x}) \mathbf{v}_k(\boldsymbol{x})^T \tag{82}$$

$$\sum_{\boldsymbol{\alpha}} B^j_{\boldsymbol{\alpha}} \mathrm{v}_k(\boldsymbol{x}) \mathrm{v}_k(\boldsymbol{x})^T = g_j(\boldsymbol{x}) \mathbf{v}_{k-1}(\boldsymbol{x}) \mathbf{v}_{k-1}(\boldsymbol{x})^T, \; \forall j = 1, \ldots, m \tag{83}$$

Thus, by Proposition 3, for appropriate sum-of-squares polynomials $(\sigma_j)_{j=0,1,\ldots,m} \subset \mathbb{R}[\boldsymbol{x}]_{k-1}$:

$$\mathrm{Tr}(\sum_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}} A^C_{\boldsymbol{\alpha}} X^C) = p_C \circ \rho^X_C(\boldsymbol{x}) \tag{84}$$

$$\mathrm{Tr}(\sum_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}} B^j_{\boldsymbol{\alpha}} Y_j) = g_j(\boldsymbol{x}) \sigma_j(\boldsymbol{x}) \tag{85}$$

$$\mathrm{Tr}(\sum_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}} A_{\boldsymbol{\alpha}} Z) = \sigma_0(\boldsymbol{x}) \tag{86}$$

This is exactly the constraint in (DSDP-CF^k) with, for convenience, $g_0 = 1$. $\quad\square$

References

1. Abramsky, S.: Relational databases and Bell's theorem. In: Tannen, V., Wong, L., Libkin, L., Fan, W., Tan, W.C., Fourman, M. (eds.) In Search of Elegance in the Theory and Practice of Computation: Essays Dedicated to Peter Buneman, Lecture Notes in Computer Science, vol. 8000, pp. 13–35. Springer (2013). https://doi.org/10.1007/978-3-642-41660-6_2
2. Abramsky, S.: Relational hidden variables and non-locality. Studia Logica **101**(2), 411–452 (2013). https://doi.org/10.1007/s11225-013-9477-4
3. Abramsky, S.: Contextual semantics: From quantum mechanics to logic, databases, constraints, and complexity. In: Dzhafarov, E., Jordan, S., Zhang, R., Cervantes, V. (eds.) Contextuality from Quantum Physics to Psychology, Advanced Series on Mathematical Psychology, vol. 6, pp. 23–50. World Scientific (2015). https://doi.org/10.1142/9789814730617_0002
4. Abramsky, S., Barbosa, R.S., Karvonen, M., Mansfield, S.: A comonadic view of simulation and quantum resources (2019). To appear in Proceedings of the 34th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS 2019)

5. Abramsky, S., Barbosa, R.S., Kishida, K., Lal, R., Mansfield, S.: Contextuality, cohomology and paradox. In: Kreutzer, S. (ed.) 24th EACSL Annual Conference on Computer Science Logic (CSL 2015), Leibniz International Proceedings in Informatics (LIPIcs), vol. 41, pp. 211–228. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik (2015). https://doi.org/10.4230/LIPIcs.CSL.2015.211

6. Abramsky, S., Barbosa, R.S., Mansfield, S.: Contextual fraction as a measure of contextuality. Phys. Rev. Lett. **119**(5), 050504 (2017). https://doi.org/10.1103/PhysRevLett.119.050504

7. Abramsky, S., Barbosa, R.S., de Silva, N., Zapata, O.: The quantum monad on relational structures. In: Larsen, K.G., Bodlaender, H.L., Raskin, J.K. (eds.) 42nd International Symposium on Mathematical Foundations of Computer Science (MFCS 2017), Leibniz International Proceedings in Informatics (LIPIcs), vol. 83, pp. 35:1–35:19. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik (2017). https://doi.org/10.4230/LIPIcs.MFCS.2017.35

8. Abramsky, S., Brandenburger, A.: The sheaf-theoretic structure of non-locality and contextuality. New J. Phys. **13**(11), 113036 (2011). https://doi.org/10.1088/1367-2630/13/11/113036

9. Abramsky, S., Gottlob, G., Kolaitis, P.G.: Robust constraint satisfaction and local hidden variables in quantum mechanics. In: Rossi, F. (ed.) 23rd International Joint Conference on Artificial Intelligence (IJCAI 2013), pp. 440–446. AAAI Press (2013)

10. Abramsky, S., Hardy, L.: Logical Bell inequalities. Phys. Rev. A **85**(6), 062114062114 (2012). https://doi.org/10.1103/PhysRevA.85.062114

11. Abramsky, S., Mansfield, S., Barbosa, R.S.: The cohomology of non-locality and contextuality. In: Jacobs, B., Selinger, P., Spitters, B. (eds.) Proceedings of 8th International Workshop on Quantum Physics and Logic (QPL 2011), Electronic Proceedings in Theoretical Computer Science, vol. 95, pp. 1–14. Open Publishing Association (2012). https://doi.org/10.4204/EPTCS.95.1

12. Abramsky, S., Sadrzadeh, M.: Semantic unification: A sheaf-theoretic approach to natural language. In: Casadio, C., Coecke, B., Moortgat, M., Scott, P. (eds.) Categories and Types in Logic, Language, and Physics: Essays dedicated to Jim Lambek on the occasion of his 90th birthday, pp. 1–13. Springer (2014). https://doi.org/10.1007/978-3-642-54789-8_1

13. Acín, A., Fritz, T., Leverrier, A., Sainz, A.B.: A combinatorial approach to nonlocality and contextuality. Commun. Math. Phys. **334**(2), 533–628 (2015). https://doi.org/10.1007/s00220-014-2260-1

14. Asadian, A., Budroni, C., Steinhoff, F.E., Rabl, P., Gühne, O.: Contextuality in phase space. Phys. Rev. Lett. **114**(25), 250403 (2015). https://doi.org/10.1103/PhysRevLett.114.250403

15. Barbosa, R.S.: On monogamy of non-locality and macroscopic averages: examples and preliminary results. In: Coecke, B., Hasuo, I., Panangaden, P. (eds.) 11th Workshop on Quantum Physics and Logic (QPL 2014), Electronic Proceedings in Theoretical Computer Science, vol. 172, pp. 36–55. Open Publishing Association (2014). https://doi.org/10.4204/EPTCS.172.4

16. Barbosa, R.S.: Contextuality in quantum mechanics and beyond. Ph.D. thesis, University of Oxford (2015)

17. Barvinok, A.: A Course in Convexity, Graduate Studies in Mathematics, vol. 54. American Mathematical Society (2002). https://doi.org/10.1090/gsm/054

18. Bell, J.S.: On the Einstein Podolsky Rosen paradox. Physics Physique Fizika **1**(3), 195 (1964). https://doi.org/10.1103/PhysicsPhysiqueFizika.1.195

19. Bell, J.S.: On the problem of hidden variables in quantum mechanics. Rev. Mod. Phys. **38**(3), 447–452 (1966). https://doi.org/10.1103/RevModPhys.38.447

20. Bermejo-Vega, J., Delfosse, N., Browne, D.E., Okay, C., Raussendorf, R.: Contextuality as a resource for models of quantum computation with qubits. Phys. Rev. Lett. **119**(12), 120505 (2017). https://doi.org/10.1103/PhysRevLett.119.120505

21. Billingsley, P.: Probability and Measure. Wiley Series in Probability and Mathematical Statistics. Wiley (1979)

22. Bogachev, V.I.: Measure theory. Springer (2007). https://doi.org/10.1007/978-3-540-34514-5

23. Brandenburger, A., Keisler, H.J.: Use of a canonical hidden-variable space in quantum mechanics. In: Coecke, B., Ong, L., Panangaden, P. (eds.) Computation, Logic, Games, and Quantum Foundations. The Many Facets of Samson Abramsky: Essays dedicated to Samson Abramsky on the occasion of his 60th birthday, pp. 1–6. Springer (2013). https://doi.org/10.1007/978-3-642-38164-5_1

24. Brandenburger, A., Yanofsky, N.: A classification of hidden-variable properties. J. Phys. A Math. Theor. **41**(42), 425302 (2008). https://doi.org/10.1088/1751-8113/41/42/425302

25. Braunstein, S.L., Van Loock, P.: Quantum information with continuous variables. Rev. Mod. Phys. **77**(2), 513 (2005). https://doi.org/10.1103/RevModPhys.77.513

26. Cabello, A., Severini, S., Winter, A.: Graph-theoretic approach to quantum correlations. Phys. Rev. Lett. **112**(4), 040401 (2014). https://doi.org/10.1103/PhysRevLett.112.040401

27. Carù, G.: Detecting contextuality: Sheaf cohomology and all vs nothing arguments. Master's thesis, University of Oxford (2015). Available at http://www.cs.ox.ac.uk/files/7608/Dissertation.pdf

28. Carù, G.: On the cohomology of contextuality. In: Duncan, R., Heunen, C. (eds.) 13th International Conference on Quantum Physics and Logic (QPL 2016), Electronic Proceedings in Theoretical Computer Science, vol. 236, pp. 21–39. Open Publishing Association (2017). https://doi.org/10.4204/EPTCS.236.2

29. Carù, G.: Towards a complete cohomology invariant for non-locality and contextuality. Preprint arXiv:1807.04203 [quant-ph] (2018)

30. Cunha, M.T.: On measures and measurements: a fibre bundle approach to contextuality. Preprint arXiv:1903.08819 [math.PR] (2019)

31. Dickson, W.M.: Quantum Chance and Non-locality: Probability and Non-locality in the Interpretations of Quantum Mechanics. Cambridge University Press (1998). https://doi.org/10.1017/CBO9780511524738

32. Duarte, C., Amaral, B.: Resource theory of contextuality for arbitrary prepare-and-measure experiments. J. Math. Phys. **59**(6), 062202 (2018). https://doi.org/10.1063/1.5018582

33. Dzhafarov, E.N., Kujala, J.V., Cervantes, V.H.: Contextuality-by-default: A brief overview of ideas, concepts, and terminology. In: Atmanspacher, H., Filk, T., Pothos, E. (eds.) 9th International Conference on Quantum Interaction (QI 2015), Lecture Notes in Computer Science, vol. 9535, pp. 12–23. Springer (2015). https://doi.org/10.1007/978-3-319-28675-4_2

34. Dzhafarov, E.N., Zhang, R., Kujala, J.: Is there contextuality in behavioural and social systems? Theme issue on Quantum probability and the mathematical modelling of decision making. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. **374**(2058), 20150099 (2016). https://doi.org/10.1098/rsta.2015.0099

35. Einstein, A., Podolsky, B., Rosen, N.: Can quantum-mechanical description of physical reality be considered complete? Phys. Rev. **47**(10), 777 (1935). https://doi.org/10.1103/PhysRev.47.777

36. Fine, A.: Hidden variables, joint probability, and the bell inequalities. Phys. Rev. Lett. **48**(5), 291 (1982). https://doi.org/10.1103/PhysRevLett.48.291

37. Folland, G.B.: Real Analysis: Modern Techniques and Their Applications. Pure and Applied Mathematics., 2nd edn. Wiley (1984) (1999)

38. Fritz, T.: Possibilistic physics (2009). https://fqxi.org/community/forum/topic/569. FQXI Essay Contest 2009

39. Giry, M.: A categorical approach to probability theory. In: Banaschewski, B. (ed.) Categorical aspects of topology and analysis. Lecture Notes in Mathematics, vol. 915, pp. 68–85. Springer (1982). https://doi.org/10.1007/BFb0092872

40. Griffiths, R.B.: Quantum measurements and contextuality. Philos. Trans. R. Soc. A **377**(2157), 20190033 (2019). https://doi.org/10.1098/rsta.2019.0033

41. Haviland, E.: On the momentum problem for distribution functions in more than one dimension. ii. Am. J. Math. **58**(1), 164–168 (1936). https://doi.org/10.2307/2371063

42. He, Q.Y., Cavalcanti, E.G., Reid, M.D., Drummond, P.D.: Bell inequalities for continuous-variable measurements. Phys. Rev. A **81**(6), 062106 (2010). https://doi.org/10.1103/PhysRevA.81.062106

43. Henaut, L., Catani, L., Browne, D.E., Mansfield, S., Pappa, A.: Tsirelson's bound and Landauer's principle in a single-system game. Phys. Rev. A **98**(6), 060302 (2018). https://doi.org/10.1103/PhysRevA.98.060302

44. Hilbert, D.: Über die darstellung definiter formen als summe von formenquadraten. Mathematische Annalen **32**(3), 342–350 (1888). https://doi.org/10.1007/BF01443605

45. Hofmann, T., Schölkopf, B., Smola, A.J.: Kernel methods in machine learning. Ann. Stat. **36**(3), 1171–1220 (2008). https://doi.org/10.1214/009053607000000677

46. Howard, M., Wallman, J., Veitch, V., Emerson, J.: Contextuality supplies the 'magic' for quantum computation. Nature **510**(7505), 351–355 (2014). https://doi.org/10.1038/nature13460

47. Jarrett, J.P.: On the physical significance of the locality conditions in the Bell arguments. Spec. Issue Found. Quant. Mech. Noûs **18**(4), 569–589 (1984). https://doi.org/10.2307/2214878

48. Kakutani, S.: Concrete representation of abstract (M)-spaces (a characterization of the space of continuous functions). Ann. Math. **42**(4), 994–1024 (1941). https://doi.org/10.2307/1968778

49. Karvonen, M.: Categories of empirical models. In: Selinger, P., Chiribella, G. (eds.) 15th International Conference on Quantum Physics and Logic (QPL 2018), Electronic Proceedings in Theoretical Computer Science, vol. 287, pp. 239–252. Open Publishing Association (2019). https://doi.org/10.4204/EPTCS.287.14

50. Ketterer, A., Laversanne-Finot, A., Aolita, L.: Continuous-variable supraquantum nonlocality. Phys. Rev. A **97**(1), 012133 (2018). https://doi.org/10.1103/PhysRevA.97.012133

51. Kishida, K.: Logic of local inference for contextuality in quantum physics and beyond. In: Chatzigiannakis, I., Mitzenmacher, M., Rabani, Y., Sangiorgi, D. (eds.) 43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016), Leibniz International Proceedings in Informatics (LIPIcs), vol. 55, pp. 113:1–113:14. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik (2016). https://doi.org/10.4230/LIPIcs.ICALP.2016.113

52. Kochen, S., Specker, E.P.: The problem of hidden variables in quantum mechanics. J. Math. Mech. **17**(1), 59–87 (1967)

53. Lasserre, J.B.: Global optimization with polynomials and the problem of moments. SIAM J. Optim. **11**(3), 796–817 (2001). https://doi.org/10.1137/S1052623400366802

54. Lasserre, J.B.: Moments, Positive Polynomials and Their Applications, Series on Optmization and Its Applications, vol. 1. Imperial College Press (2009). https://doi.org/10.1142/p665

55. Lasserre, J.B.: A new look at nonnegativity on closed sets and polynomial optimization. SIAM J. Optim. **21**(3), 864–885 (2011). https://doi.org/10.1137/100806990

56. Laurent, M.: Sums of Squares, Moment Matrices and Optimization Over Polynomials, pp. 157–270. Springer New York, New York, NY (2009). https://doi.org/10.1007/978-0-387-09686-5_7

57. Laversanne-Finot, A., Ketterer, A., Barros, M.R., Walborn, S.P., Coudreau, T., Keller, A., Milman, P.: General conditions for maximal violation of non-contextuality in discrete and continuous variables. J. Phys. A Math. Theor. **50**(15), 155304 (2017). https://doi.org/10.1088/1751-8121/aa6016

58. Luenberger, D.G.: Optimization by Vector Space Methods. Wiley (1997)

59. MacLane, S., Moerdijk, I.: Sheaves in Geometry and Logic: A First Introduction to Topos Theory. Universitext. Springer (1992). https://doi.org/10.1007/978-1-4612-0927-0

60. Mansfield, S.: Consequences and applications of the completeness of hardy's nonlocality. Phys. Rev. A (2017). https://doi.org/10.1103/PhysRevA.95.022122

61. Mansfield, S., Barbosa, R.S.: Extendability in the sheaf-theoretic approach: Construction of Bell models from Kochen–Specker models. Preprint arXiv:1402.4827 [quant-ph] (2014)

62. Mansfield, S., Fritz, T.: Hardy's non-locality paradox and possibilistic conditions for non-locality. Found. Phys. **42**(5), 709–719 (2012). https://doi.org/10.1007/s10701-012-9640-1

63. Mansfield, S., Kashefi, E.: Quantum advantage from sequential-transformation contextuality. Phys. Rev. Lett. **121**(23), 230401 (2018). https://doi.org/10.1103/PhysRevLett.121.230401

64. McKeown, G., Paris, M.G., Paternostro, M.: Testing quantum contextuality of continuous-variable states. Phys. Rev. A **83**(6), 062105 (2011). https://doi.org/10.1103/PhysRevA.83.062105

65. Okay, C., Roberts, S., Bartlett, S.D., Raussendorf, R.: Topological proofs of contextuality in quantum mechanics. Quant. Inf. Comput. **17**(13–14), 1135–1166 (2017). https://doi.org/10.26421/QIC17.13-14

66. Okay, C., Tyhurst, E., Raussendorf, R.: The cohomological and the resource-theoretic perspective on quantum contextuality: common ground through the contextual fraction. Preprint arXiv:1806.04657 [quant-ph] (2018)

67. Orbanz, P.: Probability theory ii (2011)

68. Panangaden, P.: Labelled Markov Processes. Imperial College Press (2009). https://doi.org/10.1142/p595

69. Parrilo, P.A.: Semidefinite programming relaxations for semialgebraic problems. Math. Program. **96**(2), 293–320 (2003). https://doi.org/10.1007/s10107-003-0387-5

70. Parthasarathy, K.: Probability Measures on Metric Spaces. Elsevier (1967). https://doi.org/10.1016/C2013-0-08107-8. https://linkinghub.elsevier.com/retrieve/pii/C20130081078

71. Plastino, Á.R., Cabello, A.: State-independent quantum contextuality for continuous variables. Phys. Rev. A **82**(2), 022114 (2010). https://doi.org/10.1103/PhysRevA.82.022114

72. Popescu, S., Rohrlich, D.: Quantum nonlocality as an axiom. Found. Phys. **24**(3), 379–385 (1994). https://doi.org/10.1007/BF02058098

73. Putinar, M.: Positive polynomials on compact semi-algebraic sets. Indiana Univ. Math. J. **42**(3), 969–984 (1993)

74. Raussendorf, R.: Contextuality in measurement-based quantum computation. Phys. Rev. A **88**(2), 022322 (2013). https://doi.org/10.1103/PhysRevA.88.022322

75. Raussendorf, R.: Cohomological framework for contextual quantum computations. Preprint arXiv:1602.04155 [quant-ph] (2016)

76. Raussendorf, R., Briegel, H.J.: A one-way quantum computer. Phys. Rev. Lett. **86**(22), 5188 (2001). https://doi.org/10.1103/PhysRevLett.86.5188

77. Roumen, F.: Cohomology of effect algebras. In: Duncan, R., Heunen, C. (eds.) 13th International Conference on Quantum Physics and Logic (QPL 2016), Electronic Proceedings in Theoretical Computer Science, vol. 236, pp. 174–201. Open Publishing Association(2017). https://doi.org/10.4204/EPTCS.236.12

78. Shimony, A.: Events and processes in the quantum world. In: Penrose, R., Isham, C.J. (eds.) Quantum Concepts in Space and Time, pp. 182–203. Oxford University Press (1986). https://doi.org/10.1017/CBO9781139172196.011

79. Simmons, A.W.: How (maximally) contextual is quantum mechanics? Preprint arXiv:1712.03766 [quant-ph] (2017)

80. Slofstra, W.: Tsirelson's problem and an embedding theorem for groups arising from non-local games. Preprint arXiv:1606.03140 [quant-ph] (2016)

81. Spekkens, R.W.: Contextuality for preparations, transformations, and unsharp measurements. Phys. Rev. A **71**(5), 052108 (2005). https://doi.org/10.1103/PhysRevA.71.052108

82. Su, H.Y., Chen, J.L., Wu, C., Yu, S., Oh, C.: Quantum contextuality in a one-dimensional quantum harmonic oscillator. Phys. Rev. A **85**(5), 052126 (2012). https://doi.org/10.1103/PhysRevA.85.052126

83. Vestrup, E.: The Theory of Measures and Integration. Wiley Series in Probability and Statistics. Wiley (2003)
84. Weedbrook, C., Pirandola, S., García-Patrón, R., Cerf, N.J., Ralph, T.C., Shapiro, J.H., Lloyd, S.: Gaussian quantum information. Rev. Mod. Phys. **84**(2), 621 (2012). https://doi.org/10.1103/RevModPhys.84.621
85. Wester, L.: Classical and quantum structures of computation. Ph.D. thesis, University of Oxford (2018)
86. Yoshikawa, J., Yokoyama, S., Kaji, T., Sornphiphatphong, C., Shiozawa, Y., Makino, K., Furusawa, A.: Invited article: generation of one-million-mode continuous-variable cluster state by unlimited time-domain multiplexing. APL Photon. **1**(6), 060801 (2016). https://doi.org/10.1063/1.4962732

Communicated by H-T. Yau